WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

(I1) International Publication Number:

WO 98/07864

(43) International Publication Date:

26 February 1998 (26.02.98)

(21) International Application Number:

PCT/GB97/02273

A1

(22) International Filing Date:

22 August 1997 (22.08.97)

(30) Priority Data:

9617671.4 9625996.5 23 August 1996 (23.08.96) GB 13 December 1996 (13.12.96) GB

(60) Parent Application or Grant

(63) Related by Continuation

LIS Filed on

08/782.893 (CIP) 27 December 1996 (27.12.96)

(71) Applicants (for all designated States except US): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CEN-TRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]; 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUINN,

Conrad, Padraig [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG

(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG. US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

(57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

9807864A1_I_> BNSDOCID: <WO____

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AŁ	Albania	ES	Spain	LS	Lesotho	Sì	Slovenia
AM	Armenia	FI	Finland	1.T	Lithuania	SK	Slovakia
ΑT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	, GA	Gabon	LV	l_atvia	SZ.	Swaziland
A7.	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Taükistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	1L	Israel	MR	Mauritania	UG	Uganda
BY	Belans	18	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL.	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand	2	
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ.	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	Li	Liechtenstein	SD	Sudan		
DK	Denmark	ŁK	Sri Lanka	· SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the $H_{\rm C}$ domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the $H_{\rm N}$ domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The $H_{\rm N}$ domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the $H_{\rm C}$, amino acid residues 449-871 for the $H_{\rm N}$ and residues 1-448 for the LC. Digestion with trypsin effectively degrades the $H_{\rm C}$ domain of the BoNT/A to generate a non-toxic fragment designated $LH_{\rm N}$,

BNSDOCID: <WO_____9807864A1_I_>

- 2 -

which is no longer able to bind to and enter neurons (Fig. 1). The LH_N fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

(A) clostridial neurotoxin light chain:

-a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma - membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.

(B) clostridial neurotoxin heavy chain H_N domain:

- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H_c domain.
- -a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H_N of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H_C domain.

For the purposes of the invention, the functional property or properties of the H_N of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a H_N domain or to the functions of a H_N domain are references to this property or properties. The second domain is not required to exhibit other properties of the H_N domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH₄₂₃/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

BNSDOCID: <WO_

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and H_N components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the invention described in an example below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain H_N portion or a fragment or variant of a clostridial toxin heavy chain H_N portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the H_N domain. Teachings of regions within the H_N responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the H_N domain or fragment, though it too retains the function of the H_N domain. It is conveniently obtained by insertion, deletion and/or substitution of a H_N domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a H_N domain or fragment, (iii) a Momain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

αŽ

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated $H_{\rm C}$ of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated $H_{\rm C}$ of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H_N of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H_N sequences of botulinum toxin types A, B, C₁, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic lgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endopeptidase activity of the composition will, of course, also be determined by the amount of the derivative that is present.

While it is known to treat naturally produced clostridial toxin to remove the $H_{\rm c}$ domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH_{423}/A (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH_{728}/B (SEQ ID NO: 20).

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The LH_N/A derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the H_C domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH₄₂₃/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H_C domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH_N derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The LH_N enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nomenclature,

- 13 -

LH₄₂₃/A = SEQ ID NO: 2, containing the entire L-chain and 423 amino acids of the H-chain of botulinum neurotoxin type A;

₂LH₄₂₃/A = a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;

 $_{2}L_{/2}H_{423}/A$ = a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain and the H-chain:

²L_{FXa/2}H₄₂₃/A = a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and

 $_2L_{FXa/2}H_{423}/A$ -IGF-1 = a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH₄₂₃/A;

- Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native" LH_N/A and an embodiment of the invention designated $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) in an *in vitro* peptide cleavage assay;
- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/₄H₄₂₃/A illustrating insertion of four amino acids at the N-terminus of the H_N sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H_N sequence then begins ALN...;
- Fig. 6 shows the transition region of an embodiment of the invention designated L_{FXa/3}H₄₂₃/A illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H_N will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -IGF-1, a fusion protein; the IGF-1 sequence begins at position G_{882} ;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{\text{FXa/3}}H_{423}/A\text{-CtxA14}$, a fusion protein; the C-terminal CtxA sequence begins at position Ω_{882} ;
- Fig.9 shows the C-terminal portion of the amino acid sequence of an

embodiment of the invention designated $L_{FXa/3}H_{423}/A-ZZ$, a fusion protein; the C-terminal ZZ sequence begins at position A_{890} immediately after a genenase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11 polypeptides of the invention; Fig. 10 shows LH₄₂₃/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage

sequence;

Fig. 12 shows the trypsin sensitive activation region of a polypeptide of the invention;

Fig. 13

shows Western blot analysis of recombinant LH_{107}/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH_{107}/B ; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH_{107}/B .

sites R1, R2 and R3 and a C-terminal fusion peptide

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:

Sequence

1

DNA coding for LH₄₂₃/A

2	LH ₄₂₃ /A
3	DNA coding for $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$), of which an N-terminal portion is shown in Fig. 4.
4	₂₃ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
5	DNA coding for ₂ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y), of which an N-
	terminal portion is shown in Fig.4
6	₂ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
7 .	DNA coding for native BoNT/A according to Binz et al
8	native BoNT/A according to Binz et al
9	DNA coding for L _{/4} H ₄₂₃ /A
10	L _{/4} H ₄₂₃ /A
11 .	DNA coding for $L_{FXa}/_3H_{423}/A$
12	L _{FXa} / ₃ H ₄₂₃ /A
13	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
14	L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
15	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-CtxA14
16	L _{FX3} / ₃ H ₄₂₃ /A-CtxA14
17	DNA coding for L _{FXa/3} H ₄₂₃ /A-ZZ
18	$L_{FXa/3}H_{423}/A-ZZ$
19	DNA coding for LH ₇₂₈ /B
20	LH ₇₂₈ /B
21	DNA coding for LH ₄₁₇ /B
22	LH ₄₁₇ /B
23	DNA coding for LH ₁₀₇ /B
24	LH ₁₀₇ /B
25	DNA coding for LH_{423}/A ($Q_2E,N_{26}K,A_{27}Y$)
26	$LH_{423}/A (Q_2E, N_{26}K, A_{27}Y)$
27	DNA coding for LH ₄₁₇ /B wherein the first 274 bases are
•	

WO 98/07864

PCT/GB97/02273

- 17 -

modified to have an E. coli codon bias

28

DNA coding for LH₄₁₇/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H_c) of botulinum neurotoxin type A. This recombinant product is designated the LH₄₂₃/A fragment (SEQ ID NO: 2).

Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the LH_{423}/A coding sequence was PCR amplified from total chromosomal DNA from $Clostridium\ botulinum\$ and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth Expression of the gene zi then induced by addition isopropylthio- β -D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na₂HPO₄, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH₄₂₃/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

- 19 -

The recombinant GST- LH_{423}/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH_{423}/A is recovered in the non-adsorbed material.

Construct variants

A variant of the molecule, LH_{423}/A ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH_{423}/A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A:

Two further variants of the gene sequence that have been expressed and the corresponding products purified are $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a Eco 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the Eco 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, $L_{IA}H_{423}/A$ (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of $L_{IA}H_{423}/A$ at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L_{FXa/3}H₄₂₃/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L₇₄H₄₂₃/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of $L_{FXa/3}H_{423}/A$ have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1) $L_{Fxa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2) $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3) $L_{FXa/3}H_{4Z3}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a generase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing H_N function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide

WO 98/07864 PCT/GB97/02273 - 22 -

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the H_N domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant GST- $_2$ LH $_{423}$ /A (Q_2 E, N_{26} K, A_{27} Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 μ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 μ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$, 1.8 mM KH $_2$ PO $_4$, pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

₂LH₄₂₃/A is more stable in the presence of trypsin and more active in the in vitro

WO 98/07864 PCT/GB97/02273 - 23 -

peptide cleavage assay than is 23LH423/A. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH₇₂₈/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polypeptide, LH_{417}/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

WO 98/07864 PCT/GB97/02273
- 24 -

equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native LH_{N}/A .

A gene has also been assembled coding for a variant polypeptide, LH_{107}/B (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

Construct Variants

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-₂LH₄₂₃/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH_N/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na₂HPO₄ 1.15 g/l, KH₂PO₄ 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20 μ g of GST- $_2$ LH $_{423}$ /A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH $_{\rm N}$ /A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na $_2$ HPO $_4$ anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 μ g (5X10-6 g) to 5 picograms (5X10-12 g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded

over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- $GST_{-2}LH_{423}/A$ antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-2LH₄₂₃/A antiserum.

	Botulinum Toxin/mouse														
Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)								
1	0.	4	4	4	4	4	4								
2	•	4	4	4	4	• 4	4								
3	-	4	4	4	. 4	4	4								
4	-	4	4	4	4	4	4								

TABLE 2. Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

	•	_					
Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	0	O	0	, o	2	4
. 2	-	•	-	-		0	4
з.	-	-	- .	-	-	-	4
4	-	-	-	-	-	-	4

Botulinum Toxin/mouse

Example 4

Expression of recombinant LH₁₀₇/B in E. coli.

As an exemplification of the expression of a nucleic acid coding for a LH_N of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide LH_{107}/B (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in *E. coli* BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

LH₄₂₃/A.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
 - (B) STREET: 14 Kensington Square
 - (C) CITY: London
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): W8 5HH
 - (A) NAME: FOSTER; Keith Alan
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: QUINN; Conrad Padraig
 - (B) STREET: Centre For Applied Microbiology And Research,
 Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: SHONE; Clifford Charles
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	48
Val	Asp	116	20	lyr	IIe	Lys	11e	Pro 25	Asn	Ala	GGC Gly	Gln	Met 30	Gln	Pro	96
Val	Lys	35	Pne	Lys	TIE	Hls	Asn 40	Lys	Ile	Trp	GTT Val	Ile 45	Pro	Glu	Arg	144
Asp	50	Pne	inr	Asn	Pro	55 55	Glu	Gly	Asp	Leu	AAC Asn 60	Pro	Pro	Pro	Glu	192
65	rvs	GIN	vai	Pro	70	Ser,	Tyr	Tyr	Asp	Ser 75	ACC Thr	Tyr	Leu	Ser	Thr 80	240
ASP	ASI	GIU	Lys	85	Asn	Tyr	Leu	Lys	90	Val	ACC Thr	Lys	Leu	Phe 95	Glu	288
Arg	116	lyr	100	Inr	Asp	ren	GIA	Arg 105	Met	Leu	CTG Leu	Thr	Ser 110	Ile	Val	336
Arg	GIY	11e 115	Pro	Pne	Trp	GIA	Gly 120	Ser	Thr	Ile	GAC Asp	Thr 125	Glu	Leu	Lys	384
Val	130	Asp	Thr	Asn	Cys	11e 135	Asn	Val	Ile	Gln	CCA Pro 140	Asp	Gly	Ser	Tyr	432
145	ser	GIU	GIu	Leu	150	Ļeu	Val	Ile	Ile	Gly 155	CCC Pro	Ser	Ala	Asp	Ile 160	480
iie	GIN	Pne	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	GTG Val	Leu	Asn	Leu 175	Thr	528
Arg	Asn	GIA	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	TTC Phe	Ser	Pro 190	Asp	Phe	576
Thr	Pne	195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	ACC Thr	Asn 205	Pro	Leu	Leu	624
GIY	210	GIY	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	ACC Thr 220	Leu	Ala	His	Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720

CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	Asn	ACC	AAC Asn	GCC Ala	TAC Tyr 250	Тух	GAC Glu	ATO	G AGT	GG: Gly 259	TTA / Leu	76B
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	Phe	GGI Gly	GGC Gly	CAT His	GAT S Asp 270	Ala	AAG A Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG	TAC Tyz 285	Tyr	TAC	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	Lys	TCC Ser	ATI	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	`AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	1440
ATT Ile	ACA Thr	TCT Ser	GAT Asp	ACT Thr 485	AAT Asn	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT Ile	AGT Ser 495	TTA Leu	1488
GAT Asp	TTA Leu	ATA Ile	CAA Gln 500	CAA Gln	TAT Tyr	TAT Tyr	Leu	ACC Thr 505	TTT Phe	AAT Asn	TTT Phe	GAT Asp	AAT Asn 510	GAA Glu	CCT Pro	. 153 6

GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	ÀGT Ser	GAC Asp	ATT	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	1	L 5 84
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA Arg	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT Tyr	GAG Glu	1	1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	1	L728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	3	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1	1824
CAA Gln	TTA Leu 610	GTA Val	TAT	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	:	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	:	1920
TTA Leu	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu	:	1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala	;	2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT	TTT Phe	GCA Ala 680	Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys	:	2064
Val	CTA Leu 690	Thr	GTT Val	Gln	ACA Thr	Ile	Asp	AAT Asn	GCT Ala	Leu	AGT Ser 700	Lys	AGA Arg	AAT Asn	GAA Glu		2112
AAA Lys 705	Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	Lys	TAT	ATA	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	:	2160
					GAT Asp					Lys						,	2208
				Glu	GCA Ala				Ile								2256
			Glu		A GAG			Asr					Ile				2304
		: Sei			raa 1 12a L		ı Ser					Met			ATA Ile		2352

- 33 -

AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
			GAA Glu				TAA *									2616

- (2) INFORMATION FOR SEQ ID NO: 2:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile

150

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 185 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 330 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 440 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 555 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 570 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 615 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 730 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 760 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 775 Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 810 Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 840 Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 855 Thr Phe Thr Glu Tyr Ile Lys

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2685 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2685

(xi)	SEOUENCE	DESCRIPTION:	SEO	ID	NO:	3:
------	----------	--------------	-----	----	-----	----

GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT His	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu		48
GAA Glu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn		96
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	GTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC Tyr	ATC Ile 45	AAA Lys	ATT Íle	CCA Pro	<i>:</i>	144
										TTC Phe							192
ATC Ile 65	TGG Trp	GTT Val	ATT Ile	CCG Pro	GAA Glu 70	CGC Arg	GAT Asp	ACA Thr	TTT Phe	ACG Thr 75	AAC Asn	CCG Pro	GAA Glu	GAA Glu	GGA Gly 80		240
										GTG Val							288
										AAG Lys							336
										TCC Ser							384
										CCA Pro							432
										ACT Thr 155							480
					Ser					GAA Glu							528
				Ala					Phe	GAG Glu							576
			Leu					Asn		TAC			Thr		TAC Tyr		624

ATT Ile	CGT Arg 210	TTC Phe	AGC Ser	CCA Pro	GAC Asp	TTC Phe 215	ACG Thr	TTC Phe	GGT Gly	TTC Phe	GAG Glu 220	GAG Glu	AGC Ser	CTG Leu	GAG Glu	672
GTT Val 225	GAT Asp	ACC Thr	AAC Asn	CCG Pro	CTG Leu 230	TTG Leu	GGT Gly	GCA Ala	GGC	AAG Lys 235	TTC Phe	GCA Ala	ACT Thr	GAT Asp	CCA Pro 240	720
GCG Ala	GTG Val	ACC Thr	CTG Leu	GCA Ala 245	CAC His	GAG Glu	CTG Leu	ATC Ile	CAC His 250	GCC Ala	GGT Gly	CAT His	CGT Arg	CTG Leu 255	TAT Tyr	768
GGC	ATT Ile	GCG Ala	ATT Ile 260	AAC Asn	CCG Pro	AAC Asn	CGC Arg	GTG Val 265	TTC Phe	AAG Lys	GTT Val	AAC Asn	ACC Thr 270	AAC Asn	GCC Ala	816
TAC Tyr	TAC	GAG Glu 275	ATG Met	AGT Ser	GGT Gly	TTA Leu	GAA Glu 280	GTA Val	AGC Ser	TTC Phe	GAG Glu	GAA Glu 285	CTG Leu	CGC Arg	ACG Thr	864
TTC Phe	GGT Gly 290	GGC Gly	CAT His	GAT Asp	GCG Ala	AAG Lys 295	TTT Phe	ATC Ile	GAC Asp	AGC Ser	TTG Leu 300	CAG Gln	GAG Glu	AAC Asn	GAG Glu	912
TTC Phe 305	CGT Arg	CTG Leu	TAC Tyr	TAC	TAC Tyr 310	AAC Asn	AAG Lys	TTT	AAA Lys	GAT Asp 315	ATT Ile	GCA Ala	AGT Ser	ACA Thr	CTG Leu 320	960
AAC Asn	AAG Lys	GCT Ala	AAG Lys	TCC Ser 325	ATT Ile	GTG Val	GGT Gly	ACC Thr	ACT Thr 330	GCT Ala	TCA Ser	TTA Leu	CAG Gln	TAT Tyr 335	ATG Met	1008
AAA Lys	AAT Asn	GTT Val	TTT Phe 340	AAA Lys	GAG Glu	AAA Lys	TAT Tyr	CTC Leu 345	CTA Leu	TCT Ser	GAA Glu	GAT Asp	ACA Thr 350	TCT Ser	GGA Gly	1056
AAA Lys	TTT Phe	TCG Ser 355	GTA Val	GAT Asp	AAA Lys	TTA Leu	AAA Lys 360	TTT Phe	GAT Asp	AAG Lys	TTA Leu	TAC Tyr 365	AAA Lys	ATG Met	TTA Leu	1104
ACA Thr	GAG Glu 370	ATT Ile	TAC Tyr	ACA Thr	GAG Glu	GAT Asp 375	AAT Asn	TTT Phe	GTT Val	AAG Lys	TTT Phe 380	TTT Phe	AAA Lys	GTA Val	CTT Leu	1152
Asn	Arq	Lvs	Thr	Tyr	TTG Leu 390	Asn	Phe	Asp	Lys	Ala	Val	Phe	AAG Lys	ATA Ile	AAT Asn 400	1200
ATA Ile	GTA Val	CCT Pro	AAG Lys	GTA Val 405	AAT Asn	TAC Tyr	ACA Thr	ATA Ile	TAT Tyr 410	GAT Asp	GGA Gly	TTT Phe	AAT Asn	TTA Leu 415	AGA Arg	1248
AAT Asn	ACA Thr	AAT Asn	TTA Leu 420	GCA Ala	GCA Ala	AAC Asn	TTT Phe	AAT Asn 425	GGT Gly	CAA Gln	AAT Asn	ACA Thr	GAA Glu 430	ATT Ile	AAT Asn	1296
AAT Asn	ATG Met	AAT Asn 435	TTT Phe	ACT Thr	AAA Lys	CTA Leu	AAA Lys 440	AAT Asn	TTT Phe	ACT Thr	GGA Gly	TTG Leu 445	TTT Phe	GAA Glu	TTT Phe	1344
TAT Tyr	AAG Lys 450	TTG Leu	CTA Leu	TGT Cys	GTA Val	AGA Arg 455	GGG Gly	ATA Ile	ATA Ile	ACT Thr	TCT Ser 460	AAA Lys	ACT Thr	AAA Lys	TCA Ser	1392
TTA Leu 465	GAT Asp	AAA Lys	GGA Gly	TAC Tyr	AAT Asn 470	AAG Lys	GCA Ala	TTA Leu	AAT Asn	GAT Asp 475	TTA Leu	TGT Cys	ATC Ile	AAA Lys	GTT Val 480	1440

AAT Asn	AAT Asn	TGG Trp	GAC Asp	TTG Leu 485	TTT Phe	TTT Phe	AGT Ser	CCT Pro	TCA Ser 490	GAA Glu	GAT Asp	AAT Asn	TTI Phe	ACT Thr	AAT Asn		1488
GAT A sp	CTA Leu	AAT Asn	AAA Lys 500	GGA Gly	GAA Glu	GAA Glu	ATT Ile	ACA Thr 505	TCT Ser	GAT Asp	ACT Thr	AAT Asn	ATA Ile 510	Glu	GCA Ala		1536
GCA Ala	GAA Glu	GAA Glu 515	AAT Asn	ATT Ile	AGT Ser	TTA Leu	GAT Asp 520	TTA Leu	ATA Ile	CAA Gln	CAA Gln	TAT Tyr 525	Tyr	TTA Leu	ACC Thr		1584
TTT Phe	AAT Asn 530	TTT Phe	GAT Asp	AAT Asn	GAA Glu	CCT Pro 535	GAA Glu	AAT Asn	ATT Ile	TCA Ser	ATA Ile 540	GAA Glu	AAT Asn	CTT Leu	TCA Ser		1632
545	Asp.	116	116	GIY	550	TTA Leu	GIU	ьеп	Met	Pro 555	Asn	Ile	Glu	Arg	Phe 560		1680
PIO	ASII	GIY	Lys	565	Tyr	GAG Glu	Leu	Asp	Lys 570	Tyr	Thr	Met	Phe	His 575	Tyr		1728
ьец	Ary	AIA	580	GIU	Pne	GAA Glu	HIS	585	rvs	Ser	Arg	Ile	Ala 590	Leu	Thr		1776
ASII	261	595	ASII	Glu	АТА	TTA Leu	600	Asn	Pro	Ser	Arg	Val 605	Tyr	Thr	Phe		1824
FIIC	610	261	Asp	Tyr	vai	AAG Lys 615	rys	Val	Asn	Lys	Ala 620	Thr	Glu	Ala	Ala		1872
625	FIIC	Deu	GIÀ	TIP	630	GAA Glu	Gin	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640		1920
	261	GIU	val	645	·	ACG Thr	Asp	Lys	11e 650	Ala	Asp	Ile	Thr	11e 655	Ile		1968
116	FIO	TYL	660	GIY	PIO	GCT Ala	Leu	Asn 665	Ile	Gly	Asn	Met	Leu 670	Tyr	Lys		2016
ASP	Asp	675	vai	GIY	АТА	TTA Leu	680 116	Phe	Ser	Gly	Ala	Val 685	Ile	Leu	Leu	;	2064
GIU	690	116	PIO	GIU	116	GCA Ala 695	116	Pro	Val	Leu	Gly 700	Thr	Phe	Ala	Leu		2112
GTA Val 705	TCA Ser	TAT Tyr	ATT Ile	GCG Ala	AAT Asn 710	AAG Lys	GTT Val	CTA Leu	ACC Thr	GTT Val 715	CAA Gln	ACA Thr	ATA Ile	GAT Asp	AAT Asn 720	;	2160
GCT Ala	TTA Leu	AGT Ser	AAA Lys	AGA Arg 725	AAT Asn	GAA Glu	AAA Lys	TGG Trp	GAT Asp 730	GAG Glu	GTC Val	TAT Tyr	AAA Lys	TAT Tyr 735	ATA Ile	:	2208
GTA Val	ACA Thr	AAT Asn	TGG Trp 740	TTA Leu	GCA Ala	AAG Lys	GTT Val	AAT Asn 745	ACA Thr	CAG Gln	ATT Ile	GAT Asp	CTA Leu 750	ATA Ile	AGA Arg	;	2256

AAA Lys	AAA Lys	ATG Met 755	AAA Lys	GAA Glu	GCT Ala	TTA Leu	GAA Glu 760	AAT Asn	CAA Gln	GCA Ala	GAA Glu	GCA Ala 765	ACA Thr	AAG Lys	GCT Ala	2304
ATA Ile	ATA Ile 770	AAC Asn	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn 775	CAA Gln	TAT Tyr	ACT Thr	GAG Glu	GAA Glu 780	GAG Glu	AAA Lys	AAT Asn	AAT Asn	235 2
ATT Ile 785	AAT Asn	TTT Phe	AAT Asn	ATT Ile	GAT Asp 790	GAT Asp	TTA Leu	AGT Ser	TCG Ser	AAA Lys 795	CTT Leu	AAT Asn	GAG Glu	TCT Ser	ATA Ile 800	2400
AAT Asn	AAA Lys	GCT Ala	ATG Met	ATT Ile 805	AAT Asn	ATA Ile	AAT Asn	AAA Lys	TTT Phe 810	TTG Leu	AAT Asn	CAA Gln	TGC Cys	TCT Ser 815	GTT Val	2448
TCA Ser	TAT Tyr	TTA Leu	ATG Met 820	AAT Asn	TCT Ser	ATG Met	ATC Ile	CCT Pro 825	TAT Tyr	GGT Gly	GTT Val	AAA Lys	CGG Arg 830	TTA Leu	GAA Glu	2496
				AGT Ser												2544
				TTA Leu												2592
				AGT Ser												2640
				TTA Leu 885										TAA * 895		2685

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 895 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu
1 10 15

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20 25 30 -

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro
35 40 45

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 65 70 75 80

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr 85 90 95

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys
100 105 110

Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly 180 185 His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu 215 Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala 265 Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met 325 330 Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly 345 Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu 370 375 Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn 390 Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn 425 Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe 440 Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser 455 460

PCT/GB97/02273

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 470 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 505 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 550 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 600 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 615 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 650 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 680 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 695 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 710 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 745 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 760 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 790 795 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val

WO 98/07864 PCT/GB97/02273

- 42 -

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 820 825 830

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 835 840 845

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 850 860

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 865 870 875 880

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 885 890 895

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2622 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2622
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

			TTC Phe 5									48
			ATT Ile							Gln		96
			GCT Ala									144
			TTT Phe									192
			CAG Gln									240
	_		GAG Glu 85								_	288
			TAT Tyr									336
		Gly	ATC Ile			Gly			Asp			384
	Val		GAC Asp		Cys							432

A0 Se 14	er ly	C AG	A TC	r GA	A GAN N Glu 150	r rei	AA 1 LEA 1	C CTO	C GTA	A ATO	e Il	C GG e Gl	G CC y Pr	C TC o Se	C GCG r Ala 160	
GA As	C AT	T ATO	C CAC	TTT Phe 165	Glu	TGC Cys	Lys	G AGO	TTT r Phe 170	≥ Gly	C CA	C GA s Gl	A GTO	G TT l Le 17	G AAC u Asn 5	528
CI	G AC	G CGT	0AA 1 3 Asi 180	ı Gl	TAC Tyr	GGC Gly	TCI Ser	Thi 185	Glr	TAC Tyr	C AT	r cg:	TTO Pho 190	≘ Se	C CCA r Pro	576
GA As	C TTO P Phe	ACC Thr 195	: Phe	GGT Gly	TTC Phe	GAG Glu	GAG Glu 200	Ser	CTC Leu	GAC Glu	GTT Val	GAT L Asp 209	Thi	C AA	C CCG n Pro	624
CT Le	G TT0 u Leu 210	ս Glչ	GCA Ala	GGC Gly	AAG Lys	Phe 215	Ala	ACI Thr	GAT Asp	CCA Pro	GCC Ala 220	a Val	ACC Thi	CTC	G GCA u Ala	672
CA Hi 22	s Glu	CTG Leu	ATC Ile	CAC His	GCC Ala 230	GGT Gly	CAT His	CGT Arg	CTG	TAT Tyr 235	Gly	ATI	GCC Ala	ATT	AAC Asn 240	720
CC(Pr	G AAC O Asr	CGC Arg	GTG Val	TTC Phe 245	Lys	GTT Val	AAC Asn	ACC	AAC Asn 250	Ala	TAC	TAC	GAG Glu	ATO Met 255	AGT Ser	768
GG'	r TTA / Leu	GAA Glu	GTA Val 260	Ser	TTC Phe	GAG Glu	GAA Glu	CTG Leu 265	CGC Arg	ACG Thr	TTC Phe	GGT Gly	GGC Gly 270	His	GAT Asp	816
GC(Ala	AAG Lys	Phe 275	Ile	GAC Asp	AGC Ser	TTG Leu	CAG Gln 280	GAG Glu	AAC Asn	GAG Glu	TTC Phe	CGT Arg 285	CTG Leu	TAC	TAC	864
TAC Tyr	AAC Asn 290	. Lys	TTT Phe	AAA Lys	GAT Asp	ATT Ile 295	GCA Ala	AGT Ser	ACA Thr	CTG Leu	AAC Asn 300	AAG Lys	GCT Ala	AAG Lys	TCC Ser	912
AT7 11e 305	GTG Val	GGT Gly	ACC Thr	ACT Thr	GCT Ala 310	TCA Ser	TTA Leu	CAG Gln	TAT Tyr	ATG Met 315	AAA Lys	AAT Asn	GTT Val	TTT Phe	AAA Lys 320	960
GAC Glu	AAA Lys	TAT Tyr	CTC	CTA Leu 325	TCT Ser	GAA Glu	GAT Asp	ACA Thr	TCT Ser 330	GGA Gly	AAA Lys	TTT Phe	TCG Ser	GTA Val 335	GAT Asp	1008
AAA Lys	Leu	AAA Lys	TTT Phe 340	GAT Asp	AAG Lys	TTA Leu	TAC Tyr	AAA Lys 345	ATG Met	TTA Leu	ACA Thr	GAG Glu	ATT Ile 350	TAC Tyr	ACA Thr	1056
GAG Glu	GAT Asp	AAT Asn 355	TTT Phe	GTT Val	AAG Lys	TTT Phe	TTT Phe 360	AAA Lys	GTA Val	CTT Leu	AAC Asn	AGA Arg 365	AAA Lys	ACA Thr	TAT Tyr	1104
TTG Leu	AAT Asn 370	TTT Phe	GAT Asp	AAA Lys	Ala	GTA Val 375	TTT Phe	AAG Lys	ATA Ile	Asn	ATA Ile 380	GTA Val	CCT Pro	AAG Lys	GTA Val	1152
AAT Asn 385	TAC Tyr	ACA Thr	ATA Ile	TAT Tyr	GAT Asp 390	GGA Gly	TTT Phe	AAT Asn	Leu	AGA Arg 395	AAT Asn	ACA Thr	AAT Asn	TTA Leu	GCA Ala 400	1200
GCA Ala	AAC Asn	TTT Phe	Asn	GGT Gly 405	CAA Gln	AAT Asn	ACA Thr	Glu	ATT Ile 410	AAT . Asn .	AAT Asn	ATG Met	AAT Asn	TTT Phe 415	ACT Thr	1248

										77	_					
AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	TTT Phe	ACT	GGA Gly	TTG Leu	TTT Phe 425	GAA Glu	TTT Phe	TAT Tyr	AAG Lys	TTG Leu 430	CTA Leu	TGT Cys	1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT Thr	TCT Ser	AAA Lys 440	ACT Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	AAA Lys	GGA Gly	TAC Tyr	1344
AAT Asn	AAG Lys 450	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	TGT Cys	ATC Ile	AAA Lys	GTT Val	AAT Asn 460	AAT Asn	TGG Trp	GAC Asp	TTG Leu	1392
TTT Phe 465	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 470	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 475	GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 480	1440
Glu	Glu	Ile	Thr	Ser 485	GAT Asp	Thr	Asn	Ile	Glu 490	Ala	Ala	Glu	Glu	Asn 495	Ile	1488
Ser	Leu	Asp	Leu 500	Ile	CAA Gln	Gln	Tyr	Tyr 505	Leu	Thr	Phe	Asn	Phe 510	Asp	Asn	1536
Glu	Pro	Glu 515	Asn	Ile	TCA Ser	Ile	Glu 520	Asn	Leu	Ser	Ser	Asp 525	Ile	Ile	Gly	1584
Gln	Leu 530	Glu	Leu	Met	CCT Pro	Asn 535	Ile	Glu	Arg	Phe	Pro 540	Asn	Gly	Lys	Lys	1632
Tyr 545	Glu	Leu	Asp	Lys	TAT Tyr 550	Thr	Met	Phe	His	Tyr 555	Leu	Arg	Ala	Gln	Glu 560	1680
Phe	Glu	His	Gly	Lys 565	TCT Ser	Arg	Ile	Ala	Leu 570	Thr	Asn	Ser	Val	Asn 575	Glu	1728
Ala	Leu	Leu	Asn 580	Pro	AGT Ser	Arg	_. Val	Tyr 585	Thr	Phe	Phe	Ser	Ser 590	Asp	Tyr	1776
Val	Lys	Lys 595	Val	Asn	AAA Lys	Ala	Thr 600	Glu	Ala	Ala	Met	Phe 605	Leu	Gly	Trp	1824
Val	Glu 610	Gln	Leu	Val	TAT Tyr	Asp 615	Phe	Thr	Asp	Glu	Thr 620	Ser	Glu	Val	Ser	1872
Thr 625	Thr	Asp	Lys	Ile	GCG Ala 630	Asp	Ile	Thr	Ile	Ile 635	Ile	Pro	Tyr	Ile	Gly 640	1920
Pro	Ala	Leu	Asn	11e 645	GGT Gly	Asn	Met	Leu	Tyr 650	Lys	Asp	Asp	Phe	Val 655	Gly	1968
Ala	Leu	Ile	Phe 660	Ser	GGA Gly	Ala	Val	11e 665	Leu	Leu	Glu	Phe	11e 670	Pro	Glu	2016
ATT Ile	GCA Ala	ATA Ile 675	Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 680	Phe	GCA Ala	CTT Leu	GTA Val	TCA Ser 685	TAT Tyr	ATT Ile	GCG Ala	2064

AAT Asn	AAG Lys 690	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 695	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 700	Leu	AGT Ser	AAA Lys	AGA Arg		2112
AAT Asn 705	GAA Glu	ÄAA Lys	TGG Trp	GAT Asp	GAG Glu 710	GTC Val	TAT Tyr	AAA .Lys	TAT Tyr	ATA Ile 715	GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 720		2160
GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 725	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 730	AGA Arg	AAA Lys	AAA Lys	ATG Met	AAA Lys 735	GAA Glu		2208
GCT Ala	TTA Leu	GAA Glu	AAT Asn 740	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 745	AAG Lys	GCT Ala	ATA Ile	ATA Ile	AAC Asn 750	TAT Tyr	CAG Gln		2256
TAT Tyr	AAT Asn	CAA Gln 755	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 760	AAA Lys	AAT Asn	AAT Asn	ATT Ile	AAT Asn 765	TTT Phe	AAT Asn	ATT Ile		2304
						CTT Leu 775											2352
						AAT Asn											2400
						GTT Val										÷	2448
						AAG Lys											2496
						TTA Leu										-	2544
						CTT Leu 855											2592
TTA Leu 865	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 870	TAT Tyr	ATT Ile	AAG Lys	TAA *								2622

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 874 amino acids (B) TYPE: amino acid

 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val

Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro

WO 98/07864 - 46 - PCT/GB97/02273

Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro 60 Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala 150 Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro 185 Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 200 Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 250 Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser 295 Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys-Phe Ser Val Asp 330 Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 345 Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 390

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr 405 Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tvr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu 455 Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly 470 Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile 485 490 Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys 535 Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu 570 Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser 615 Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly 630 635 Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu 665 Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg 695 Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu 715 Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln

WO 98/07864 PCT/GB97/02273

									•	- 48	-				
Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn	Ile
Asp	Asp 770	Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile .
Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn B00
Ser	Met	Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser
Leu	Lys	Asp	Ala 820	Leu	Leu	Lys	Tyr	Ile 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu
Ile	Gly	Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser
Thr	Asp 850	Ile	Pro	Phe	Gln	Leu 855	Ser	Lys	Tyr	Val	Asp 860	Asn	Gln	Arg	Leu
Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	*						
(2)	INF	ORMA!	LION	FOR	SEQ	ID I	NO :	7:							
	(i)	() ()	A) L B) T C) S	ENGT: YPE : TRAN	HARAC H: 20 nuc: DEDNI OGY:	613 leic ESS:	base aci dou	pai: d	rs						
	(ii) MO	LECU:	LE T	YPE:	DNA	(ge	nomi	c)						
	(ix	()		AME/	KEY: ION:		613								

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

	TTT Phe								48
	ATT Ile								. 96
	GCT Ala 35					 	 	 	144
 -	TTT Phe		_	_		 	 	 	192
	CAA Gln								240
	GAA Glu	-							288

7 70/0 /004	4.0
	- 49

AGA Arg	ATT Ile	TAT Tyr	TCA Ser 100	ACT Thr	GAT Asp	CTT Leu	GGA Gly	AGA Arg 105	ATG Met	TTG Leu	TTA Leu	ACA Thr	TCA Ser 110	ATA Ile	GTA Val	336
AGG Arg	GGA Gly	ATA Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGA Gly 120	AGT Ser	ACA Thr	ATA Ile	GAT Asp	ACA Thr 125	GAA Glu	TTA Leu	AAA Lys	384
GTT Val	ATT Ile 130	GAT Asp	ACT Thr	AAT Asn	TGT Cys	ATT Ile 135	AAT Asn	GTG Val	ATA Ile	CAA Gln	CCA Pro 140	GAT Asp	GGT Gly	AGT Ser	TAT Tyr	432
AGA Arg 145	TCA Ser	GAA Glu	GAA Glu	CTT Leu	AAT Asn 150	CTA Leu	GTA Val	ATA Ile	ATA Ile	GGA Gly 155	CCC Pro	TCA Ser	GCT Ala	GAT Asp	ATT Ile 160	480
ATA Ile	CAG Gln	TTT Phe	GAA Glu	TGT Cys 165	AAA Lys	AGC Ser	TTT Phe	GGA Gly	CAT His 170	GAA Glu	GTT Val	TTG Leu	AAT Asn	CTT Leu 175	ACG Thr	528
CGA Arg	AAT Asn	GGT Gly	TAT Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAA Gln	TAC Tyr 185	ATT Ile	AGA Arg	TTT Phe	AGC Ser	CCA Pro 190	GAT Asp	TTT Phe	576
ACA Thr	TTT Phe	GGT Gly 195	TTT Phe	GAG Glú	GAG Glu	TCA Ser	CTT Leu 200	GAA Glu	GTT Val	GAT Asp	ACA Thr	AAT Asn 205	CCT Pro	CTT Leu	TTA Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAA Lys	TTT Phe	GCT Ala	ACA Thr 215	GAT Asp	CCA Pro	GCA Ala	GTA Val	ACA Thr 220	TTA Leu	GCA Ala	CAT His	GAA Glu	672
CTT Leu 225	ATA Ile	CAT His	GCT Ala	GGA Gly	CAT His 230	AGA. Arg	TTA Leu	TAT Tyr	GGA Gly	ATA Ile 235	GCA Ala	ATT Ile	AAT Asn	CCA Pro	AAT Asn 240	720
AGG Arg	GTT Val	TTT Phe	AAA Lys	GTA Val 245	AAT Asn	ACT Thr	AAT Asn	GCC Ala	TAT Tyr 250	TAT Tyr	GAA Glu	ATG Met	Ser	GGG Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTT Phe 260	GAG Glu	GAA Glu	CTT Leu	AGA Arg	ACA Thr 265	TTT Phe	GGG Gly	GGA Gly	CAT His	GAT Asp 270	GCA Ala	AAG Lys	816
Phe	Ile	Asp	Ser	Leu	CAG Gln	Glu	Asn	Glu	Phe	Arg	Leu	Tyr	TAT Tyr	TAT Tyr	AAT Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATA Ile	GCA Ala	AGT Ser 295	ACA Thr	CTT Leu	AAT Asn	AAA Lys	GCT Ala 300	AAA Lys	TCA Ser	ATA Ile	GTA Val	912
GGT Gly 305	ACT Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104

TTT Phe	GAT Asp 370	Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	11	52
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	12	00
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	12	48
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	12	96
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	134	44
ATA	Leu 450	Asn	Asp	Leu	Cys	11e 455	Lys	Val	Asn	Asn	TGG Trp 460	Asp	Leu	Phe	Phe	139	92
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	144	4 0
lle	Thr	Ser	Asp	Thr 485	Asn	Ile	Glu	Ala	Ala 490	Glu	GAA Glu	Asn	Ile	Ser 495	Leu	148	88
Asp	Leu	IIe	500	Gin	Tyr	Tyr	Leu	Thr 505	Phe	Asn	TTT Phe	Asp	Asn 510	Glu	Pro	153	36
GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT Ser	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	156	34
Glu	Leu 530	Met	Pro	Asn	Ile	-Glu 535	Arg	Phe	Pro	Asn	GGA Gly 540	Lys	Lys	Tyr	Glu	163	32
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	168	30
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	172	8
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	177	16
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GÁA Glu	182	4
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	187	2
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	192	0

WO 98/07864	W	O'	98/	'O '	78	64
-------------	---	----	-----	-------------	----	----

TTA Leu	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu		1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala		2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys		2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu		2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT. Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	•	2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu		2208
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAĠ Gln 750	TAT Tyr	AAT Asn	•	2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp		2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile		2352
AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800		2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys		2448
GAT Asp	GCA Ala	TTA Leu	Tra Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly		2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp		2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser		2592
					ATT Ile 870												2613

(2) INFORMATION FOR SEQ ID NO: 8:

⁽i) SEQUENCE CHARACTERISTICS:

⁽A) LENGTH: 871 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- 52 -

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 40 Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 330 325

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 470 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 520 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 570 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 600 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 630 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 650 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys

- 54 -

 Val
 Leu
 Thr
 Val
 Gln
 Thr
 Ile
 Asp
 Asn
 Ala
 Leu
 Ser
 Lys
 Asp
 Asn
 Glu

 Lys
 Trp
 Asp
 Glu
 Val
 Tyr
 Lys
 Tyr
 Ile
 Val
 Thr
 Asn
 Trp
 Leu
 Ala
 Lys
 Lys
 Lys
 Met
 Lys
 Glu
 Ala
 Leu
 Tyr
 Asn
 Lys
 Met
 Lys
 Glu
 Ala
 Leu
 Tyr
 Asn
 Lys
 Asn
 Tyr
 Glu
 Ala
 Ile
 Asn
 Lys
 Asn
 Asn
 Tyr
 Asn
 Asn
 Asn
 Tyr
 Asn
 Asn
 Asn
 Tyr
 Asn
 Asn
 Asn
 Tyr
 Asn
 Asn
 Ile
 Asn
 Tyr
 Asn
 Ile
 Asn
 Tyr
 Asn
 Ile
 Asn
 Tyr
 Leu
 Asn
 Ile
 Asn
 Ile
 Asn
 Ile
 Asn
 Asn
 Ile
 Asn

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2628 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2628
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

			CAG Gln						48
			AAA Lys						96
	Phe		CAT His						144
 -			GAA Glu 55	Glu	 Asp	Leu	 	 	 192

GCA Ala 65	AAG Lys	CAG Gln	GTG Val	Pro	GTI Val 70	Ser	TAC	TAC Tyl	GAT Asp	TC/ Sei	Th	C TA	r CTC	G AG	C ACA r Thr 80	240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	Asn	TAC	CTG Leu	Lys	GGA Gly 90	r Val	ACC L Th:	C AA	A TTI	A TTO 1 Pho 9	C GAG e Glu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	Thr	GAC Asp	CTG Leu	GGC	Arg 105	, Met	CTC Lev	CTO	ACC Thi	TC/ Ser 110	Ile	GTC Val	336
CGC	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG	GGT	GGC Gly 120	Ser	ACC Thr	ATI	GAC Asp	C ACC	Glu	TTO Lev	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro) Asp	GGI Gly	AGC Ser	TAC	432
AGA Arg 145	Ser	GAA Glu	GAA Glu	CTT	AAC Asn 150	CTC Leu	GTA Val	ATC	ATC Ile	GGG Gly 155	CCC	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC	CAC His 170	GAA Glu	GTG Val	Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT	TTC	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg.	GTG Val	TTC Phe	ÄAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT . Tyr !	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT (GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	Asp	AAA Lys 335	TTA Leu	1008

											="						
AAA Lys	TTT	GAT Asp	AAG Lys 340		TAC	AAA Lys	ATG Met	TTA Leu 345	inr	GAG Glu	ATI	TAC	ACA Thr	Glı	GAT Asp	1056	6
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	Ten	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	Tyr	TTO	AAT Asn	1104	4
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	Lys	GTA Val	AAT Asn	TAC	1152	2
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200)
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248	3
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296	\$
GGG	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344	
AGC Ser	GCT Ala 450	GAT Asp	GGG Gly	GCA Ala	TTA Leu	AAT Asn 455	GAT Asp	TTA Leu	TGT Cys	ATC Ile	AAA Lys 460	GTT Val	AAT Asn	AAT Asn	TGG Trp	1392	:
GAC Asp 465	TTG . Leu	TTT Phe	TTT Phe	AGT Ser	CCT Pro 470	TCA Ser	GAA Glu	GAT Asp	AAT Asn	TTT Phe 475	ACT Thr	AAT Asn	GAT Asp	CTA Leu	AAT Asn 480	1440	
AAA Lys	GGA Gly	GAA Glu	GAA Glu	ATT Ile 485	ACA Thr	TCT Ser	GAT Asp	ACT Thr	AAT Asn 490	ATA Ile	GAA Glu	GCA Ala	GCA Ala	GAA Glu 495	GAA Glu	1488	
AAT Asn	ATT Ile	AGT Ser	TTA Leu 500	GAT Asp	TTA Leu	ATA Ile	CAA Gln	CAA Gln 505	TAT Tyr	TAT Tyr	TTA Leu	ACC Thr	TTT Phe 510	AAT Asn	TTT Phe	1536	
GAT Asp	AAT Asn	GAA Glu 515	PIU	GAA Glu	AAT Asn	ATT Ile	TCA Ser 520	тте	GAA Glu	AAT A sn	CTT Leu	TCA Ser 525	AGT Ser	GAC Asp	ATT Ile	1584	
ATA Ile	GGC Gly 530	CAA Gln	TTA Leu	GAA Glu	CTT Leu	ATG Met 535	CCT Pro	AAT Asn	ATA Ile	GAA Glu	AGA Arg 540	TTT Phe	CCT Pro	AAT Asn	GGA Gly	1632	
AAA Lys 545	AAG Lys	TAT Tyr	GAG Glu	TTA Leu	GAT Asp 550	AAA Lys	TAT Tyr	ACT Thr	ATG Met	TTC Phe 555	CAT His	TAT Tyr	CTT Leu	CGT Arg	GCT Ala 560	1680	
CAA Gln	GAA Glu	TTT Phe	GAA Glu	CAT His 565	GGT Gly	AAA Lys	TCT Ser	AGG Arg	ATT Ile 570	GCT Ala	TTA Leu	ACA Thr	AAT Asn	TCT Ser 575	GTT Val	1728	
AAC Asn	GAA Glu	GCA Ala	TTA Leu 580	TTA Leu	AAT Asn	CCT Pro	AGT Ser	CGT Arg 585	GTT Val	TAT Tyr	ACA Thr	TTT Phe	TTT Phe 590	TCT Ser	TCA Ser	1776	
GAC Asp	TAT Tyr	GTA Val 595	AAG Lys	AAA Lys	GTT Val	AAT Asn	AAA Lys 600	GCT Ala	ACG Thr	GAG Glu	GCA Ala	GCT Ala 605	ATG Met	TTT Phe	TTA Leu	1824	

GGC TGG Gly Trp 610	GTA GAA (Val Glu (Jan Dea v	TA TAT G al Tyr A 15	AT TTT A sp Phe T	CC GAT GAR hr Asp Glu 620	A ACT AGC GA Thr Ser GI	AA 1872 lu
625	,	630	ie Ala A	sp lie T	hr Ile Ile 35	ATT CCA TA Ile Pro Ty 64	0
ile Giy i	6	45	re Gry As	650	eu Tyr Lys	GAT GAT TI Asp Asp Ph 655	e
vai Giy A	660	TO THE B	66 66	sa val 1.	le Leu Leu	GAA TTT AT Glu Phe Il 670	e
6	75	ie rio va	680 680	y inr pr	e Ala Leu 685	GTA TCA TA Val Ser Ty	r
690	зи шув У	69	val Gi	n Thr II	e Asp Asn 700	GCT TTA AG Ala Leu Se	r
705	51. OZU <i>D</i> .	710	p din va	71 71 71	5 Tyr Ile	GTA ACA AAT Val Thr Asr 720))
. IIp bed A	Ta Lys va	25	r Gin II	e Asp Le 730	u Ile Arg	AAA AAA ATO Lys Lys Met 735	
nys Gru A	740	lu Asn GI	n Ala Git 745	Ala Th	r Lys Ala	ATA ATA AAO Ile Ile Asn 750	•
75	71 ASH GI	n lyr in	760	i Giu Lys	s Asn Asn 1 765	ATT AAT TTT Ile Asn Phe	
770	sp Asp Le	77!	Lys Leu	Asn Glu	Ser Ile 1 780	AAT AAA GCT Asn Lys Ala	
785	in lie As	790	e Leu Asn	Gln Cys 795	Ser Val S	CCA TAT TTA Ser Tyr Leu 800	2400
Met Asii Se	80:	e pro Tyr 5	GIY Val	Lys Arg 810	Leu Glu A	SAT TTT GAT Asp Phe Asp 815	2448
. WIR SET TE	820	o Ala Leu	Beu Lys 825	Tyr Ile	Tyr Asp A 8	AT AGA GGA sn Arg Gly	2496
ACT TTA AT Thr Leu Il 83	e GIY GII	ı val Asp	840	Lys Asp	Lys Val A 845	sn Asn Thr	2544
CTT AGT AC Leu Ser Th	A GAT ATA r Asp Ile	CCT TTT Pro Phe 855	Gin Leu	TCC AAA Ser Lys	TAC GTA G Tyr Val A 860	AT AAT CAA sp Asn Gln	2592
AGA TTA TTA Arg Leu Leu 865	A TCT ACA 1 Ser Thr	TTT ACT Phe Thr 870	GAA TAT Glu Tyr	ATT AAG Ile Lys 875	*		2628

- (2) INFORMATION FOR SEQ ID NO: 10:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 876 amino acids
 - (B) TYPE: amino acid(D) TOPOLOGY: linear

 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 155 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val

295

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 440 Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 505 Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile 520 Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val 565 Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 615 Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 650

Val	Gly	Ala	Leu 660	Ile	Phe	Ser	Gly	Ala 665	Val	Ile	Leu	Leu	Glu 670	Phe	Il
Pro	Glu	Ile 675	Ala	Ile	Pro	Val	Leu 680	Gly	Thr	Phe	Ala	Leu 685	Val	Ser	Ту
	0,50					095					Asp 700				
		-			, 10					/15	Tyr				720
				123					730		Ile			735	
			740					/45			Lys		750		
•		, 35		•			760				Asn	765			
	,,,					//5					Ser 780				
,05					790					795	Ser				800
				803					810		Leu			815	
			620					825			Tyr		830		_
		033					840				Lys	845			
	050					022					Tyr 860	Val	Asp	Asn	Gln
Arg	Leu	Leu	Ser	Thr	Phe	Thr	Glu	Tyr	Ile	Lys	*				

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2637 base pairs
 - (B) TYPE: nucleic acid

870

- (C) STRANDEDNESS: double (D) TOPOLOGY: linear
- (ii) MOLECULE_TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2637
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 30

GTG Val	AAG Lys	GCT Ala 35	Phe	AAC Lys	ATT	CAT His	AAC Asn 40	Lys	A ATO	TGC Tr	G GT	r AT	e Pro	G GA	A CGC	144	4
GAT Asp	ACA Thr 50	Phe	ACG Thr	AAC Asr	CCG Pro	GAA Glu 55	Glu	GG# Gly	A GAC / Asp	TTO Lev	AAC ASI 60	n Pro	G CCC	G CCC	G GAA G Glu	192	2
GCA Ala 65	Lys	CAG Gln	GTG Val	Pro	GTI Val 70	Ser	TAC	TAC	GAT Asp	TCA Ser 75	Thi	TA	CTC Lev	G AGO	ACA Thr	240)
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	Asn	TAC	CTG Leu	AAC Lys	GGA Gly 90	' Val	ACC Thr	Lys	TT#	TTO Phe 95	GAG Glu	288	3
CGT	ATT	TAT	TCC Ser 100	Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	Met	Leu	Leu	ACC Thr	Ser 110	Ile	GTC Val	336	;
CGC	GGA Gly	ATC Ile 115	CCA Pro	TTT	TGG	GGT Gly	GGC Gly 120	AGT Ser	ACC	ATT	GAC Asp	ACG Thr 125	Glu	TTG Leu	AAG Lys	384	r -
GTT Val	ATT Ile 130	GAC	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC. Asn	GTG Val	ATC	CAA Gln	CCA Pro 140	Asp	GGT	AGC Ser	TAC	432	
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480	
ATC Ile	CAG Gln	TTT Phe	Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528	
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT	TTC Phe	AGC Ser	CCA Pro 190	Asp	TTC - Phe	576	
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624	
GCT	GCA Ala 210	Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672 ·	
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	. 720	
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768	
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816	
T TT Phe	Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG .	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864	
Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	Ala	AGT : Ser ' 295	ACA Thr	CTG Leu	AAC Asn	Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912	

G0 G1 30	Y	ACC Thr	ACT Thr	GCT	TCA	TTA	CAG	TAT	ATG	AAA	AAT	GTT	TTT	AAA	GAG	AAA	960
					261	310	Gln	Tyr	Met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320	300
TA Ty	T T	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
A.F Ly	VA /S	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
A.F	AT sn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TT Pł	Je .	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
AC Th 38	ır	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
T1 Pł	TT ne	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
A. Ly	AA /s	TAA NaA	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
G(GG ly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
AT II	le	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
As	AT sn 55	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GJ As	AT Sp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	Glu	Glu	Ile	Thr	Ser	Asp	ACT Thr	Asn	Ile	Glu	Ala	1488
G(CA la	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
T;	rr he	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	CCT Pro	GAA Glu 520	TAA Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
A(GT er	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632
P	CT ro 45	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT Tyr 550	GAG Glu	TTA Leu	GÁT Asp	AAA Lys	TAT Tyr 555	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560	1680
L.	TT	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728

AA' Ası	T TC n Se	T G1 r Va	T AA il As 58	in GI	A GO	A TT a Le	A TI u Le	CA AA u As 58	in Pr	T AG	T CO	ST G1	ıl Ty	AT A	CA hr	TTT Phe		1776
TT: Phe	T TC ≥ Se	T TC r Se 59	r As	C TA	T GI	A AA 1 Ly	G AA s Ly 60	s Va	T AA l As	T AA n Ly	A GC s Al	T AC a Th 60	r Gl	AG G	CA la	GCT Ala	•	1824
AT(Met	TT Ph	e re	A GG u Gl	C TG y Tr	G GT p Va	A GA 1 Gl: 61:	ı Gl	A TT n Le	A GT u Va	A TA 1 Ty	T GA r As 62	p Ph	T AC e Th	C GA	AT Sp	GAA Glu	:	1872
ACT Thr 625	Se.	C GA	A GT. u Va	A AG 1 Se	T AC r Th 63	T ACC	G GA	T AA p Ly	A AT s Il	T GC e Al. 63	a As	T AT.	A AC e Th	T AT	.e	ATT Ile 640	:	1920
ATI Ile	Pro	A TA	T AT	A GG e Gl	y Pr	T GCT	TT	A AA' u Asi	T AT	e Gl	T AA' Y Asi	T ATO	G TT	A TA u Ty 65	T	AAA Lys		1968
GAT Asp	GAT Asp	TT Pho	r GT Va.	r GT	y Ala	r TTA a Lei	1116	∍ Phe	≥ Se:	r Gli	7 Ala	r GT a Val	F AT	e Le	G u	TTA Leu	2	016
GAA Glu	TTI Phe	116 675	∍ Pro	GA(AT:	GCA Ala	ATA Ile	Pro	GT)	TTA Leu	A GGT A Gly	Thi 685	Phe	r gc ≥ Al	A (a]	CTT Leu	2	064
GTA Val	TCA Ser 690	TY	ATI	GCC Ala	AAT ASI	Lys 695	Val	CTA Leu	ACC Thr	GTI Val	CAA Gln 700	1 Thr	ATA Ile	A GA	T A	AAT Asn	2	112
705	Leu	Ser	. Lys	Arg	710		Lys	Trp	Asp	715	Val	Туг	Lys	Ty	r] 7	le 720	2	160
vai	Inr	ASI	Trp	725	Ala	AAG Lys	Val	Asn	730	Gln	Ile	Asp	Leu	735	⊋ A	rg	2 :	208
Lys	rys	Met	140	GIu	Ala	TTA Leu	Glu	Asn 745	Gln	Ala	Glu	Ala	Thr 750	Lys	A	la	2:	256
lle	iie	755	TYT	Gln	Tyr	AAT Asn	Gln 760	Tyr	Thr	Glu	Glu	Glu 765	Lys	Asn	A	sn	23	304
IIe	770	Phe	Asn	Ile	Asp	GAT Asp 775	Leu	Ser	Ser	Lys	Leu 780	Asn	Glu	Ser	I.	le	23	352
785	rys	Ala	Met	lle	790	ATA Ile	Asn	Lys	Phe	Leu 795	Asn	Gln	Cys	Ser	V:	al 00	24	00
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GZ GZ	AA Lu	24	48
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	G# As	AT Sp	24	96
AAT A	Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GIA	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	G1 Va	T	25	44

- 64 -

AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 2592 850 855 GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 2637

- (2) INFORMATION FOR SEQ ID NO: 12:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 879 amino acids(B) TYPE: amino acid

 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12: Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 105 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 155 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn _ _ 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 345 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 375 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 440 Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 535 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 570 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 585 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 630 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 665 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 695 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 855 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870

- (2) INFORMATION FOR SEQ ID NO: 13:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2862 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2862
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATG Met 1	GIn	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	Lys	GAC Asp	CC1	GT)	A AA l As:	C GGT n Gly		48
GTT Val	GAC Asp	ATT	GCC Ala 20	TAC	ATC Ile	AAA Lys	ATT Ile	Pro 25	Asn	GCC Ala	GGC	CAC Glr	ATO Met	Gl	G CCG		96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT	CAT	AAC Asn 40	Lys	ATC Ile	TGG	GTT Val	ATI Ile 45	Pro	GAZ Glu	A CGC 1 Arg		144
GAT Asp	ACA Thr 50	Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	Pro	CCC Pro	CCC Pro	GAA Glu		192
GCA Ala 65	AAG Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC	TAC	GAT Asp	TCA Ser 75	Thr	TAT	Leu	AGC Ser	ACA Thr 80		240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC	AAA Lys	TTA Leu	Phe 95	GAG Glu		288
CGT Arg	ATT Ile	TAT	TCC Ser 100	ACT	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC	GTC Val		336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys		384
Val	11e 130	Asp	Thr	Asn	Cys	Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser			432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160		480
Ile	Gln	Phe	Glu	TGC Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr		528
Arg	Asn	Gly	Tyr 180	GGC Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe		576
Thr	Phe	Gly 195	Phe	GAG Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu		624
Gly	Ala 210	Gly	Lys	TTC Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu		672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	-	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	·	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	His.	GAT Asp 270	GCG Ala	AAG Lys		816

TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
lyr	reu	Leu	ser	325	Asp	ACA Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gl'n	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	CCT Pro	GAA Glu 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 640 625 625 630 ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 655 GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 660 GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 660 675 GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 GCT TTA AGT AAA AGA AAT GAA AAA TCG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 GTA ACA AAT TGG TTA GCA AGG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730 AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT GAT AAT ACT GGA GAA GAA AAA TAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 ATT AAT TTT AAT ATT GAT GAT TAA ATA AAT TCG AAA CTT AAT GAG TCT ATA ASN Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val ASN TAAA GCT ATG ATT AAT AAT AAA TAAA TTT TTG AAT CAC TCT GTT ASN Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val ASN TABA GCT ATG ATT AAT AAT AAA TAAA TTT TTG AAT CAC TCT GTT ASN Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val ASN TABA GCT ATG ATT AAT AAT AAA TAAA TTT TTG ASN CAC TCT GTT ASN Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val ASN TATA TTT AAT ATT AAT AATA AAT AAA TATA TTT TTG AAT CAA TCG TCT GTT ASN Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val ASN TATA TTT AAT ATT AAT AATA AAT AAA TTT TTG AAT CAA TCA TCT TTT ASN CAT TCT TTT ASN CAT TCT TTT ASN CAT TCT TTT ASN CAC TCT TTT TCT TCT TTT TTT ATT TTT ATT TTT ATT TTT T																	
ACT TCT GTT AAC GAA GCA TAA TAA AAA AAT GCG GAT AAT ACC GAT GAT CAT TAT AAA GAA GAT TAT TAAAT CCT AGT CGT GTT TAT ACC GAT GAA AAA GAA AAA GGT TAA TAA AAA GCT ACC GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Ser Val Asn Grant Car ACC GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Ser Car Car Gac GAC GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Ser Car Car Gac GAC GCT Phe Ser Ser Asp Tyr Val Glu Gin Leu Val Tyr Asp Phe Thr Asp Glu Gin Leu Val Tyr Asp Phe Thr Asp Glu Gin Leu Val Tyr Asp Phe Thr Asp Glu Gin Car Gac	Pr	o Asi	r GG n Gl	A AA y Ly:	A AAG s Ly	s Ty:	r Glu	I TT)	A GA	T AA p Ly	s Ty:	r Th	T AT	G TI t Ph	C CA	is Tyr	:
ASIN SET VAI ASIN GIU AIA LEU LEU ASIN PRO SET ARG VAI TYT THE Phe 580 585 585 585 585 585 586 ASIN TIT TOT TOA GAC TAT GTA AAA AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val ASIN Lys Ala Thr Glu Ala Ala 650 600 600 600 600 600 600 600 600 600	CT.	CG7	GC' Ala	T CAI a Gli	n GII	1 Pue	T GAA	A CAT	GG' Gl	y Ly:	s Sei	r Ag	G AT g Il	T GC e Al	a Le	u Thr	1728
Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 605 595 600 600 ATG TIT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620 620 620 620 620 620 620 620 620 620	AA? Asi	TCI Ser	GT: Val	l Ası	ı Glı	A GCA 1 Ala	A TTA Leu	TTA	Ası	n Pro	r AG1 Sez	CG Ar	T GT g Va	l Ty	r Th	A TTT ir Phe	1776
Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 615 615 625 ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 665 GTA TCA TAT ATT GCG AAT AGG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 695 GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 720 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730 AAA AAA ATG AAA AGA GCT TTA GAA AAT CAA CAA GAA GAG AAA AAG AAG ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730 AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GAA GAA GAG AAA AAT AAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT TY AND TATA AGA AGA GCT TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT AAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT TY AND TATA AGA CAA GAG GAA GAG AAA AAT AAT TY GAT TATA ATT CAA TAT ACT GAG GAA CTT AAT GAG TCT ATA ATT AAT CAA TAT ACT GAG GAA CTT AAT GAG TCT ATA ATT AAT AAT CAA TAT ACT GAG GAA CTT AAT GAG TCT ATA ATT AAT TAT GAT GAT TAT AAT CAA TAT ACT GAG GAA CTT AAT GAG TCT ATA ATT AAT TAT AAT TTA ATT GAT GAT TATA AAT TATA AGT TTA AGT TATA AAT AAT	TT1 Phe	TCI Ser	Sei	c Asp	TAT Tyr	T GTA	A AAG L Lys	Lys	Va]	raa 1 lasi	AAA Lys	GC" Ala	a Th:	r Gl	G GC u Al	A GCT a Ala	1824
Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 640 ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA ASP ASP Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 666 GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 670 GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 GCT TTA AGT AAA AGA AAT GAA AAA TCG GAT GAG GTC TAT AAA TAT ATA AGA 690 GCT TTA AGT AAA AGA AAT GAA AAA TCG GAT GAG GTC TAT AAA TAT ATA AGA ACA ATA GAT ATA TTP Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 715 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG GTT GAT CTA ATA AGA ATA GAT ATA TTP Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 735 AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT CAT GAT TA ACT GAG GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 755 ATT AAT TTT AAT ATT GAT GAT TA ACT TA ACT GAG GAA GAA AAA AAT AAT TAT ATT GAT GA	ATC Met	Phe	Let	A GGC u Gly	TGC Trp	GTA Val	. Glu	Gln	TTA Leu	A GTA ı Val	TAI Tyr	Ası	Phe	r AC	C GA r As	T GAA p Glu	1872
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA ASP ASP Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 670 GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA ACA AAT TYR Leu Ala Lys Trp Asp Glu Val Tyr Lys Tyr Ile 710 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 AAAA AAAA ATG AAA GGT ATA GAT CAA TAT AAT CAA GCA GAA GCA ACA ATA GAG CT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAA GAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 760 ATT AAT TTT AAT ATT GAT GAT TAA ATA AAA TTT TTG GAA CTT AAT GAG CTT TAAT ATT TATA ATT AAT TTT AAT ATT GAT GA	Thr	Ser	GA# Glu	A GTA ı Val	AGT Ser	Thr	Thr	GAT Asp	AAA Lys	ATT Ile	Ala	Ası	T ATA	A AC	r AT	e Ile	1920
Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ATT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680 GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA ALA Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA ATA GAT ATA TATA ATA CAL CAG ATT GAT CTA ATA AGA ATA TATA ATA CAS TAT THE ASP TYP Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 735 AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA TAT ACT GAG GAA GAA AAA AAT THR Lys Glu Ala Leu Glu Asn Gln Ala Glu Lys Asn Asn 740 ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT ATA TILe Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 ATT AAT TTT AAT ATT GAT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA ATA ACT TATA TTA ATT ACT GAT GAT TTA ATT TTA ATT GAT GAT TATA ATA	ATI	CCA Pro	TAI	ATA Ile	Gly	Pro	GCT Ala	TTA Leu	AAT Asn	Ile	Gly	AA1 Asr	ATC Met	E TT	ı Ty	r Lys	1968
GIU Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp 715 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 735 AAA AAA ATA AAG GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT AAT ALI Lle Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 765 ATT AAT TTT AAT ATT GAT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA ASP Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 AAA AAA GCT ATG ATT AAT ATA AAT AAA TATA AAA TATA TTT TTG AAT CAA TGC TCT GTT ASN Lys Ala Met Ile Asn 1le Asn Lys Phe Leu Asn Gln Cys Ser Val 795 GTA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA CGT TTA AAT CAAT TTA AGT CTA TTA AGT CCT TAT GGT GTT AAA CGG TTA AAT CAA CGC TTA CAA CGC	GAT Asp	GAT Asp	TTT Phe	Val	Gly	GCT Ala	TTA Leu	ATA Ile	Phe	Ser	GGA Gly	GCT	GTI Val	. Ile	Le	G TTA u Leu	2016
Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 720 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 735 AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn 760 ATT AAT TTT AAT ATT GAT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT ASN Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 CCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT GAT GAT TTA AGT TTA AGT GAT TTA AGT CCT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT GAT TTA ATT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT GGT GTT AAT CAA TTA TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT TTA ATG AAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA 26 CTT TAT GGT GTT AAA CGG TTA GAA C	GAA Glu	TTT Phe	Ile	Pro	GAG Glu	ATT Ile	GCA Ala	Ile	CCT Pro	GTA Val	TTA Leu	GGT Gly	Thr	Phe	GC/ Ala	A CTT	2064
Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 720 GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 735 AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAA AAT AAT ASN Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 760 ATT AAT TTT AAT ATT GAT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA 1le Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 CTCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu	GTA Val	Ser	TAT	ATT	GCG Ala	AAT Asn	Lys	GTT Val	CTA Leu	ACC	GTT Val	Gln	Thr	ATA Ile	GAT Asp	AAT Asn	2112
Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 735 AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 ATT AAT TTT AAT ATT GAT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 AAT AAA GCT ATG ATT AAT ATA AAA AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 795 TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu	Ala	TTA Leu	AGT Ser	AAA Lys	AGA Arg	Asn	GAA Glu	AAA Lys	TGG Trp	GAT Asp	Glu	GTC Val	TAT Tyr	AAA Lys	TAT	Ile	2160
Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn 755 ATT AAT TTT AAT ATT GAT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 790 TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu	GTA Val	ACA Thr	AAT Asn	TGG Trp	Leu	GCA Ala	AAG Lys	GTT Val	AAT Asn	Thr	CAG Gln	ATT Ile	GAT Asp	CTA Leu	Ile	Arg	2208
Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 765 ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu	AAA Lys	AAA Lys	ATG Met	Lys	GAA Glu	GCT Ala	TTA Leu	GAA Glu	Asn	CAA Gln	GCA Ala	GAA Glu	GCA Ala	Thr	AAG Lys	GCT Ala	2256
Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu	ATA Ile	ATA Ile	Asn	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	Gln	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	Glu	AAA Lys	AAT Asn	AAT Asn	2304
Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800 TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu	ATT Ile	Asn	TTT Phe	AAT Asn	ATT Ile	GAT Asp	Asp	TTA Leu	AGT Ser	TCG Ser	AAA Lys	Leu	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352
Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu	Asn	AAA Lys	GCT Ala	ATG Met	Ile	Asn	ATA Ile	AAT . Asn	AAA Lys	Phe	Leu	AAT Asn	CAA Gln	TGC Cys	TCT Ser	Val	2400
	TCA Ser	TAT Tyr	TTA Leu	ATG Met	Asn	TCT . Ser	ATG . Met	ATC Ile	Pro	Tyr	GGT Gly	GTT Val	AAA Lys	CGG Arg	Leu	GAA Glu	2448

GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp	2496
AAT A	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val	2544
AAT A	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	2640
CCT (GGA Gly	CCG Pro	GAG Glu	ACG Thr 885	CTC Leu	TGC Cys	GGG Gly	GCT Ala	GAG Glu 890	CTG Leu	GTG Val	GAT Asp	GCT Ala	CTT Leu 895	CAG Gln	2688
TTC (GTG Val	TGT Cys	GGA Gly 900	GAC Asp	AGG Arg	GGC Gly	TTT Phe	TAT Tyr 905	TTC Phe	AAC Asn	AAG Lys	CCC Pro	ACA Thr 910	GGG Gly	TAT Tyr	2736
GGC Gly	TCC Ser	AGC Ser 915	AGT Ser	CGG Arg	AGG Arg	GCG Ala	CCT Pro 920	CAG Gln	ACA Thr	GGT Gly	ATC Ile	GTG Val 925	GAT Asp	GAG Glu	TGC Cys	2784
TGC '	TTC Phe 930	CGG Arg	AGC Ser	TGT Cys	GAT Asp	CTA Leu 935	AGG Arg	AGG Arg	CTG Leu	GAG Glu	ATG Met 940	TAT Tyr	TGC Cys	GCA Ala	CCC Pro	2832
CTC I Leu I 945	AAG Lys	CCT Pro	GCC Ala	AAG Lys	TCA Ser 950	GCT Ala	GAA Glu	GCT Ala	TAG	-						2862

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 954 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

Arg	Gly	Ile 115		Phe	Trp	Gly	Gly 120		Thr	Ile	e Asp	Thr 125		Leu	Ly
Val	Ile 130		Thr	Asn	Cys	Ile 135		Val	Ile	Gln	Pro 140		Gly	Ser	Ту
Arg 145		Glu	Glu	Leu	Asn 150		Val	lle	Ile	Gly 155		Ser	Ala	Asp	11 16
Ile	Gln	Phe	Glu	Cys 165		Ser	Phe	Gly	His 170		Val	. Leu	Asn	Leu 175	
Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185		Arg	Phe	Ser	Pro 190	-	Ph
Thr	Phe	Gly 195		Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205		Leu	Le
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215		Pro	Ala	Val	Thr 220		Alà	His	Gl
Leu 225	Ile	His	Aļa	Gly	His 230			Tyr	Gly	Ile 235		. Ile	Asn	Pro	As:
Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Lei
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Ly
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Ası
-	290			Ile		295					300	_			
305				Ser	310					315					320
Tyr	Leu	Leu	Ser	Glu 325	Asp	Thr	Ser	Gly	Lys. 330	Phe	Ser	Val	Asp	Lys 335	Let
Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr. 350		Asp
		355	·	Phe			360				-	365			
	370			Val		375					380				
385				Gly	390					395					400
		_		Asn 405					410					415	
-			420	Gly				425					430		
•		435	•	Ser			440					445			_
Ile	Glu 450	Gly	Arg	Cys	Авр	Gly 455	Ala	Leu	Asn	Asp	Leu 460	Cys	Ile	Lys	Val

					4.0		Ser	•		4/5					480
				.03			Ile		490					495	
			300				Asp	505					510		
			,				Glu 520					525			
		•				222	Glu				540				
					330		Leu			555					560
				202			His		570					575	
			.500				Leu	202	• •				590		
							Lys 600					605			
						013	Gln				620				
023					630		Asp			635					640
				043		•	Leu		650					655	
			,000				Ile	665					670		
		073					Ile 680				•	685			
	0,00					695	Val				700				
,03					110		Lys			715					720
				125			Val		730					735	
Lys	Lys	Met	Lys 740	Glu	Ala	Leu	Glu	Asn 745	Gln	Ala	Glu	Ala	Thr 750	Lÿs	Ala
Ile	Ile	Asn 755	Tyr	Gln	Tyr	Asn	Gln 760	Tyr	Thr	Glu	Glu	Glu 765	Lys	Asn	Asn
Ile	Asn 770	Phe	Asn	Ile	Asp	Asp 775	Leu	Ser	Ser	Lys	Leu 780	Asn	Glu	Ser	Ile
Asn 785	Lys	Ala	Met	Ile	Asn 790	Ile	Asn	Lys	Phe	Leu 795	Asn	Gln	Cys	Ser	Val 800
Ser	Tyr	Leu	Met	Asn 805	Ser	Met	Ile	Pro	Tyr 810	Gly	Val	Lys	Arg	Leu 815	Glu

WO 98/07864 PCT/GB97/02273

- 73 -

										, ,						
Asp	Phe	Asp	Ala 820	Ser	Leu	Lys	Asp	Ala 825	Leu	Leu	Lys	Tyr	11e		Asp	•
Asn	Arg	Gly 835	Thr	Leu	Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845	_	Lys	. Val	
Asn	Asn 850	Thr	Leu	Ser	Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860	Ser	Lys	Tyr	Val	
Asp 865		Gln	Arg	Leu	Leu 870	Ser	Thr	Phe	Thr	Glu 875		Ile	Lys	Ser	Arg 880	
Pro	Gly	Pro	Glu	Thr 885		Cys	Gly	Ala	Glu 890	Leu	Val	Asp	Ala	Leu 895	Gln	•
Phe	Val	Cys	Gly 900	Asp	Arg	Gly	Phe	Tyr 905	Phe	Asn	Lys	Pro	Thr 910	Gly	Tyr	•
Gly	Ser	Ser 915	Ser	Arg	Arg	Ala	Pro 920	Gln	Thr	Gly	Ile	Val 925	Asp	Glu	Cys	
Cys	Phe 930	Arg	Seŗ	Cys		Leu 935		Arg	Leu	Glu	Met 940	Tyr	Cys 	Ala	Pro	
Leu 945	Lys	Pro	Ala	Lys	Ser 950	Ala	Glu	Ala	, *							
(2)	INFO	ORMAT	CION	FOR	SEQ	ID N	10: 3	15:								
		MOL FEA	ECULATURE	TRANI POLO E TY : ME/I	nucledne DEDNE DGY: (PE: CEY:	ESS: line DNA CDS	doul ear (ġer	ole	:)				·			
	(xi)	SEQ	UENC	E DE	SCRI	PTIC	N: 5	EQ I	D NC): 15	5 :					
ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	4.8
GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	AAC Asn	GCC Ala	GGC	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT Ile	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	TTT.	ACG . Thr .	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
GCA Ala 65	AAG Lys	CAG (Gln	GTG Val	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
		GAG :														288

CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104

Phe	GAT Asp 370) Lys	A GCO S Ala	C GT/ a Val	A TT:	T AAC E Lys 375	s Ile	A AA:	T ATA	A GTA	A CCT	o Ly	G GT s Va	A AA l As	T TAC	1152
ACA Thr 385	: Ile	TATE TYPE	GA:	r GG/ o Gly	Y Phe 390	Asr	TTA Leu	A AGA	A AAT	ACA Thr 395	Ası	r TT.	A GC. u Al	A GC a Al	A AAC a Asn 400	1200
TTT	'AAT Asr	GGI Gly	CAZ Glr	A AAI A Asr 405	Thr	A GAA	ATI Ile	AA] Asi	AAT Asn 410	Met	AA] Ası	TTT	T AC	T AA T Ly 41	A CTA s Leu S	1248
AAA Lys	AAT Asn	TT1 Phe	ACT Thr 420	: Gly	TTC Leu	TTT Phe	GAA Glu	TTI Phe 425	Tyr	AAG Lys	Leu	CT/	TG: 1 Cys 430	s Va	A AGA l Arg	1296
GGG Gly	ATA Ile	Ile 435	Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	Ser	TTA Leu	GAT Asp	AAA Lys	GGF Gly 445	Туз	C AA'	T AAG n Lys	1344
ATC Ile	GAA Glu 450	Gly	CGT Arg	TGC Cys	Asp	GGG Gly 455	Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	Cys	`ATC	Lys	A GTT s Val	1392
Asn 465	Asn	Trp	Asp	Leu	Phe 470	Phe	Ser	Pro	Ser	Glu 475	Asp	Asn	Phe	Thi	AAT Asn 480	1440
Asp	Leu	Asn	Lys	Gly 485	Glu	Glu	Ile	Thr	Ser 490	Asp	Thr	Asn	Ile	495		1488
Ala	Glu	Glu	Asn 500	Ile	Ser	Leu	Asp	Leu 505	Ile	Gln	Gln	Tyr	Tyr 510	Leu	ACC Thr	1536
Phe	Asn	Phe 515	Asp	Asn	Glu	Pro	Glu 520	Asn	Ile	Ser	Ile	Glu 525	Asn	Leu	TCA Ser	1584
Ser	Asp 530	He	Ile	Gly	Gln	Leu 535	Glu	Leu	ATG Met	Pro	Asn 540	Ile	Glu	Arg	Phe	1632
Pro 545	Asn	Gly	Lys	Lys	Tyr 550	Glu	Leu	Asp	AAA Lys	Tyr 555	Thr	Met	Phe	His	Tyr 560	1680
Leu	Arg	Ala	Gln	Glu 565	Phe	Glu	His	Gly	AAA Lys 570	Ser	Arg	Ile	Ala	Leu 575	Thr	1728
Asn	Ser	Val	Asn 580	Glu	Ala	Leu	Leu	Asn 585	CCT Pro	Ser	Arg	Val	Tyr 590	Thr	Phe	1776
Phe '	Ser	Ser 595	Asp .	Tyr	Val	Lys	Lys 600	Val	AAT Asn	Lys .	Ala	Thr 605	Glu	Ala	Ala	1824
Met	Phe 610	Leu	Gly	Trp	Val	Glu 615	Gln .	Leu	GTA ' Val '	Tyr	Asp 620	Phe	Thr	Asp	Glu	1872
ACT Thr 625	AGC Ser	GAA Glu	GTA Val	Ser	ACT . Thr '	ACG (GAT A	AAA . Lys	ATT (GCG (Ala 1 635	GAT . Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640	1920

ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 645	CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	AAA Lys	1968
GAT Asp	GAT Asp	TTT Phe	GTA Val 660	GGT Gly	GCT Ala	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	CTG Leu	TTA Leu	2016
GAA Glu	TTT Phe	ATA Ile 675	CCA Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT Leui	2064
Val	Ser 690	Tyr	Ile	Ala	AAT Asn	Lys 695	Val	Leu	Thr	Val	Gln 700	Thr	Ile	Asp	Asn	2112
Ala 705	Leu	Ser	Lys	Arg	AAT Asn 710	Glu	Lys	Trp	Asp	Glu 715	Val	Tyr	Lys	Tyr	Ile 720	2160
Val	Thr	Asn	Trp	725	GCA Ala	Lys	Val	Asn	730	Gln	Ile	Asp	Leu	11e 735	Arg	2208
Lys	Lys	Met	140	Glu	GCT Ala	Leu	Glu	Asn 745	Gln	Ala	Glu	Ala	Thr 750	Lys	Ala	2256
Ile	Ile	755	Tyr	Gln	TAT Tyr	Asn	760	Tyr	Thr	Glu	Glu	Glu 765	Lys	Asn	Asn	2304
Ile	770	Phe	Asn	Ile	GAT Asp	775	Leu	Ser	Ser	Lys	Leu 780	Asn	Glu	Ser	Ile	2352
785	Lys	Ala	Met	Ile	AAT Asn 790	Ile	Asn	Lys	Phe	Leu 795	Asn	Gln	Cys	Ser	Val 800	2400
Ser	Tyr	Leu	Met	Asn 805	TCT Ser	Met	Ile	Pro	Tyr 810	Gly	Val	Lys	Arg	Leu 815	Glu	2448
Asp	Phe	Asp	Ala 820	Ser	CTT Leu	Lys	Asp	Ala 825	Leu	Leu	Lys	Tyr	Ile 830	Tyr	Asp	2496
Asn	Arg	Gly 835	Thr	Leu	ATT Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845	Asp	Lys	Val	2544
Asn	Asn 850	Thr	Leu	Ser	ACA Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860	Ser	Lys	Tyr	Val	2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	2640
CCT Pro	CAA Gln	TCT Ser	AAA Lys	GTT Val 885	AAA Lys	AGA Arg	CAA Gln	ATA Ile	TTT Phe 890	TCA Ser	GGC Gly	TAT Tyr	CAA Gln	TCT Ser 895	GAT Asp	2688
ATT Ile	GAT Asp	ACA Thr	CAT His 900	AAT Asn	AGA Arg	ATT Ile	AAG Lys	GAT Asp 905	GAA Glu	TTA Leu	TGA					2724

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 908 amino acids

 - (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 290

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 455 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 585 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 635 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645

WO 98/07864 PCT/GB97/02273

- 79 -

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 775

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 810

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 825

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 870

Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp

Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3042 base pairs

 - (B) TYPE: nucleic acid(C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..3042
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 10

48

Va	T GAG l Asp	Ile	GCC Ala 20	TAC	Ile	AAA Lys	ATT Ile	Pro 25	AAC Asn	GCC Ala	GGC	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GT Va	G AAG l Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
ĠA As	T ACA p Tha 50	Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
Al	A AAG a Lys 5	G CAG	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
GA As	C AAG	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CG Ar	T AT	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CG	c GGZ g Gl	A ATC	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
Va	130	Asp	ACT	Asn	Cys	11e 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr	432
Ar 14	g Se:	c Glu	GAA Glu	Leu	Asn 150	Leu	Val	Ile	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160	480
AT Il	C CAG	TTT n Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CG	T AA g Asi	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
AC Th	G TTO	GGT Gly 195	Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
G)	T GC. y Al. 21	a Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC	GAG Glu	672
CT Le 22	u Il	C CAC	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CC Az	SC GT Tg Va	3 TTC l Phe	: AAG : Lys	GTT Val 245	Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GZ GZ	VA GT Lu Va	A AGO 1 Ser	Phe 260	Glu	GAA Glu	CTG Leu	CGC	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
T?	TT AT ne Il	C GAC e Asr 275	AGC Ser	TTG	CAG Gln	GAG Glu	AAC Asn 280	Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864

AAG Lys	TTT Phe 290	Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	Thr	CTG	AAC Asn	AAC Lys	GCT Ala 300	Lys	TCC Ser	ATI	GTG Val		912
GGT Gly 305	Thr	ACT Thr	GCT Ala	TCA Ser	Leu 310	Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	Val	TTI Phe	AAA Lys	GAG Glu	AAA Lys 320		960
TAT Tyr	CTC	CTA Leu	TCT Ser	GAA Glu 325	Asp	ACA Thr	TCT	GGA Gly	AAA Lys 330	Phe	TCG Ser	GTA Val	GAT Asp	Lys 335	TTA Leu		1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	Leu	TAC	AAA Lys	ATG Met	TTA Leu 345	Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	Glu	GAT Asp		1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	Tyr	TTG Leu	AAT Asn	,	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr		1152
ACA Thr 385	ATA Ile	TAT	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	- AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400		1200
TTT Phe	AAT Asn	GCT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu		1248
A AA Lys	AAT Asn	TTT	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg		1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys		1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val		1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480		1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala		1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr		1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	CCT Pro	GAA Glu 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser		1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe		1632
CCT Pro 545	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT Tyr 550	GAG Glu	TTA Leu	GAT Asp	AAA Lys	TAT Tyr 555	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560		1680

CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728
AAT Asn	TCT Ser	GTT Val	AAC Asn 580	GAA Glu	GCA Ala	TTA Leu	TTA Leu	AAT Asn 585	CCT Pro	AGT Ser	CGT Arg	GTT Val	TAT Tyr 590	ACA Thr	TTT Phe	1776
												ACG Thr 605				1824
ATG Met	TTT Phe 610	TTA Leu	GGC	TGG Trp	GTA Val	GAA Glu 615	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 620	TTT Phe	ACC Thr	GAT Asp	GAA Glú	1872
ACT Thr 625	AGC Ser	GAA Glu	GTA Val	AGT Ser	ACT Thr 630	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640	1920
												ATG Met				1968
												GTT Val				2016
Glu	P'ne	Ile 675	Pro	Glu	Ile	Ala	Ile 680	Pro	Val	Leu	Gly	ACT Thr 685	Phe	Ala	Leu	2064
												ACA Thr				2112
												TAT				2160
					Ala							GAT Asp			Arg	2208
				Glu					Gln			GCA Ala				2256
			Tyr					Туг					Lys		AAT Asn	2304
		Phe					Leu					Asn			ATA Ile	2352
	Lys					Ile					Asr				Val 800	2400
					. Sei					Gly					GAA Glu	2448
				a Se					a Lei					Ty:	TAD 1	2496

								GAT Asp		 2544
								AAA Lys		2592
								AAG Lys		2640
								GAA Glu		2688
								GAG Glu 910	TTA Leu	2736
		Asn			Gln			ATC Ile		2784
								GAA Glu		2832
								TTC Phe		2880
 	_							AAC Asn		2928
								GAC Asp 990		2976
					Ala			GAT Asp		3024
CCG Pro 1010	Lys		TAG *							3042

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1014 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 150 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 185 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 380 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 600 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 615 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 650 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala-Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 695 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 710 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 745

- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 860
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 875 880
- Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895
- Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910
- His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925
- Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940
- Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960
- Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975
- Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990
- Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000
- Ala Pro Lys Val Asp * 1010
- (2) INFORMATION FOR SEQ ID NO: 19:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3509 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..3509
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ATO Me	g cci t Pro	A GT	T AC.	A ATA	A AA' e As: 5	T AA? n Asi	r TT	r AA' e Asi	T TA	r Ası	r GA' n As	r cc	r Ar	T GA e As 1	T AAT p Asn 5	4 B
AA: Asi	n Asi	r AT	r AT e Ile 20	e Met	S ATO	G GAC	CC1	CC Pro 25	o Phe	r GCC e Ala	AGA Arg	A GGT	Th:	r Gl	G AGA y Arg	. 96
TA:	TAT	Lys Lys	s Ala	r TTT	C AAA	A ATO	ACA Thr	: Ası	CGT Arg	T ATT	TG(ATA Ile 45	: Ile	A CC	G GAA O Glu	144
AG/ Arg	TAT TYT 50	Thi	r TTT	GGA Gly	TAT	AAA Lys	Pro	GAC Glu	GAT Asp	TTI Phe	AAT Asr 60	Lys	AG1 Se1	TÇ(GGT Gly	192
ATT Ile	: Phe	AAT Asr	r AGA n Arg	GAI Asp	GTT Val 70	. Cys	GAA Glu	TAT	TAI	GAT Asp 75	Pro	GAT Asp	TAC	TT?	AAT Asn 80	240
ACI Thr	`AAI Asn	GAT Asp	Lys	AAG Lys 85	Asn	ATA Ile	TTT Phe	TTA	CAA Glm 90	Thr	ATG Met	ATC Ile	AAG Lys	Leu 95	TTT Phe	288
AAT Asn	AGA Arg	ATC Ile	Lys 100	Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	Met	ATT Ile	336
ATA Ile	AAT Asn	GGT Gly 115	lle	CCT Pro	TAT	CTT	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	384
TTT Phe	AAC Asn 130	Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn	432
CCA Pro 145	Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160	480
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATÀ Ile	GAT Asp	ATA Ile 175	GGT Gly	528
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	576
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro	672
GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA Glu	CTT Leu	ATA Ile	CAT His	GTT Val 235	TTA Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240	720
GGC Gly	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT . Ala	Ile	CAG Gln 265	GCA Ala	GAA (Glu (GAA Glu	Leu	TAT Tyr 270	ACA Thr	TTT Phe	816

GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632

AC. Th: 54!	r Phe	r cc	T CT	A GA: u Asj	r AT. p Il	e Arç	A GAT J Asp	r ATA	A AG: ⊇ Se:	T TT r Let 55!	u Th	A TC r Sé	T TC r Se	A TI r Ph	T GAT e Asp 560	1680
GA: Asj	r GC/ p Ala	A TT.	A TT u Lei	A TT: u Phe 565	e Se	r Aac	AAA Lys	A GTT	TA1	r Sei	A TT	T TT e Ph	T TC e Se:	T AT r Me 57	G GAT t Asp 5	1728
TA:	r ATT	Ly:	A ACT s Thi 580	r Ala	AA. A Asi	r AAA n Lys	GTC Val	GT/ Val 585	Glu	A GCA 1 Ala	A GG	A TT	A TT	e Al	A GGT a Gly	1776
TG(Tr	GTO Val	E AAI Lys 595	s Glr	ATA	A GTZ Val	A AAT L Asn	GAT Asp 600	Phe	GTA Val	ATC Ile	GA/	A GC: 1 Ala 609	a Ası	r AA h Ly:	A AGC s Ser	1824
AAT Asr	ACT Thr 610	Met	G GAT	AAA Lys	ATI Ile	GCA Ala 615	Asp	ATA Ile	TCI Ser	CTA	ATT 1 116	· Val	CCT L Pro	TA:	T ATA	1872
GGA Gly 625	Leu	GCI Ala	TTA Leu	TAA A	GTA Val 630	Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	Lys	A GGZ Glý	A AAT / Ast	TTT	GAA Glu 640	1920
AAT Asn	GCT Ala	TTT Phe	GAG Glu	Ile 645	Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	Leu	CTA Leu	GAA Glu	TTI Phe	116 655	CCA Pro	1968
GAA Glu	Leu	TTA	ATA Ile 660	Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	Phe	TTA Leu	TTA Leu	GAA Glu	TCA Ser 670	Tyr	ATT	2016
GAC Asp	AAT Asn	AAA Lys 675	Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	Leu	ACI Thr	AAA Lys	2064
AGA Arg	AAT Asn 690	GAA Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	ATG Met	TAC Tyr	GGA Gly	TTA Leu	ATA Ile 700	GTA Val	GCG Ala	CAA Gln	TGG	2112
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720	2160
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr	2208
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp	2256
TTT Phe	AAT Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	AAA Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile	2304
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	TTT Phe	ATA Ile 775	AAT Asn	GGA Gly	TGT Cys	TCT Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met	2352
AAA Lys 785	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA . Leu .	AAT Asn	Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448

ጥፐ ር	ΔΤΤ	CCA	ACT	GCA	GNA	ጥአጥ	C2.2									
Leu	Ile	Gly	Ser 820	Ala	Glu	Tyr	GAA	Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC	TTG Leu	2496
AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
CTA Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	AAT Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser	GAA Glu	ATT Ile 860	TTA Leu	AAT Asn	AAT Asn	ATT Ile	2592
ATC Ile 865	TTA Leu	AAT Asn	TTA Leu	AGA Arg	TAT Tyr 870	AAG Lys	GAT Asp	AAT Asn	AAT Asn	TTA Leu 875	ATA Ile	GAT Asp	TTA Leu	TCA Ser	GGA Gly 880	2640
TAT Tyr	GGG Gly	GCA Ala	AAG Lys	GTA Val 885	GAG Glu	GTA Val	TAT Tyr	GAT Asp	GGA Gly 890	GTC Val	GAG Glu	CTT Leu	AAT Asn	GAT Asp 895	AAA Lys	2688
AAT Asn	CAA Gln	TTT Phe	AAA Lys 900	TTA Leu	ACT Thr	AGT Ser	TCA Ser	GCA Ala 905	AAT Asn	AGT Ser	AAG Lys	ATT Ile	AGA Arg 910	GTG Val	ACT Thr	2736
CAA Gln	AÄT Asn	CAG Gln 915	AAT Asn	ATC Ile	ATA Ile	TTT Phe	AAT Asn 920	AGT Ser	GTG Val	TTC Phe	CTT Lėu	GAT Asp 925	TTT Phe	AGC Ser	GTT Val	2784
AGC Ser	TTT Phe 930	TGG	ATA Ile	AGA Arg	ATA Ile	CCT Pro 935	AAA Lys	TAT Tyr	AAG Lys	AAT Asn	GAT Asp 940	GGT Gly	ATA Ile	CAA Gln	AAT Asn	2832
TAT Tyr 945	ATT Ile	CAT His	AAT Asn	GAA Glu	TAT Tyr 950	ACA Thr	ATA Ile	ATT Ile	AAT Asn	TGT Cys 955	ATG Met	AAA Lys	AAT Asn	AAT Asn	TCG Ser 960	2880
GGC Gly	TGG Trp	AAA Lys	ATA Ile	TCT Ser 965	ATT Ile	AGG Arg	GGT Gly	AAT Asn	AGG Arg 970	ATA Ile	ATA Ile	TGG Trp	ACT Thr	TTA Leu 975	ATT Ile	2928
GAT Asp	ATA Ile	AAT Asn	GGA Gly 980	AAA Lys	ACC Thr	AAA Lys	TCG Ser	GTA Val 985	TTT Phe	TTT Phe	GAA Glu	TAT Tyr	AAC Asn 990	ATA Ile	AGA Arg	2976
GAA Glu	GAT Asp	ATA Ile 995	TCA Ser	GIu	Tyr	ATA Ile	Asn	Arg	Trp	Phe	Phe	Val	Thr	ATT Ile	ACT Thr	3024
AAT Asn	AAT Asn 1010	Leu	AAT Asn	AAC Asn	GCT Ala	AAA Lys 1019	Ile	TAT Tyr	ATT Ile	AAT Asn	GGT Gly 1020	Lys	CTA Leu	GAA Glu	TCA Ser	3072
AAT Asn 1025	Thr	GAT Asp	ATT Ile	AAA Lys	GAT Asp 1030	ATA Ile	AGA Arg	GAA Glu	GTT Val	ATT Ile 1035	Ala	AAT Asn	GGT Gly	GAA Glu	ATA Ile 1040	3120
ATA Ile	TTT Phe	AAA Lys	TTA Leu	GAT Asp 104	GTA	GAT Asp	ATA Ile	GAT Asp	AGA Arg 1050	Thr	CAA Gln	TTT Phe	ATT Ile	TGG Trp 1055	Met	3168
AAA Lys	TAT Tyr	TTC Phe	AGT Ser 1060	Ile	TTT Phe	AAT Asn	ACG Thr	GAA Glu 106	Leu	AGT Ser	CAA Gln	TCA Ser	AAT Asn 1070	Ile	GAA Glu	3216
GAA Glu	AGA Arg	TAT Tyr 107	Lys	ATT Ile	CAA Gln	TCA Ser	TAT Tyr 1080	Ser	GAA Glu	TAT Tyr	TTA Leu	AAA Lys 1089	Asp	TTT Phe	TGG Trp	3264

	AAT Asn 1090	Pro					Lys					Phe				331:
AAT Asn 1105	AAA Lys	AAT Asn	TCA Ser	TAT Tyr	ATT Ile 1110	Lys	CTA Leu	AAG Lys	AAA Lys	GAT Asp 1115	Ser	CCT Pro	GTA Val	GGT Gly	GAA Glu 1120	3360
ATT Ile	TTA Leu	ACA Thr	CGT Arg	AGC Ser 1125	Lys	TAT Tyr	AAT Asn	CAA Gln	AAT Asn 1130	Ser	AAA Lys	TAT Tyr	ATA Ile	AAT Asn 1139	Tyr	3406
AGA Arg	GAT Asp	TTA Leu	TAT Tyr 1140	Ile	GGA Gly	GAA Glu	AAA Lys	TTT Phe 1145	Ile	ATA Ile	AGA Arg	AGA Arg	AAG Lys 1150	Ser	AAT Asn	3456
	CAA Gln		Ile					Val					Tyr			3504
CTA Leu	GA								-							3509

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1169 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 120 Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 135

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly

			100					102					190	Met	
		173					200					205		Gln	
Asn	Lys 210	Gly	Ala	Ser	Ile	Phe 215	Asn	Arg	Arg	Gly	Tyr 220	Phe	Ser	Asp	Pro
223					230					235				Leu	240
				245					250					Lys 255	
•			260					265					270	Thr	
		2/5					280					285	•	Ser	
	290					295					300	•	•	Leu	
303			•		310					315				Ile	320
				325					330					Glu 335	_
			340					345					350	Ser	
		333					360					365		Ile	
	3 / 0					375					380			Ile	
365					390					395				Asn	400
			•	405					.410					Ala 415	
			420				,	425					430	Val	
		433					440					445		Ile	•
	450					455					460			Phe	
465					470					475				Ser	480
				485					490					Thr 495	
			500					505					510	Leu	
Аsр	Phe	Asn 515	Val	Asp	Val	Pro	Val 520	Tyr	Glu	Lys	Gln	Pro 525	Ala	Ile	Lys

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 550 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Mét Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 600 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 650 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 760 Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 775 Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr **B**05 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly

BNEDOCID: NO

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 890 Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 905 Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 935 Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 950 Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 985 Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 1000 Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1015 Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1065 Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1080

Gly Ash Pro Leu Mer Tur Ash Lyr Gly Tyr Tyr Mor Dbo Ash Ala Gly

Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100

Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120

Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135

Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150

Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165

Leu

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- 95 -

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

	(X1)	350)()LIV		JU C. (.					J. -	•••	•					
ATG Met 1	CCA Pro	GTT Val	ACA Thr	ATA Ile 5	AAT Asn	AAT Asn	TTT Phe	AAT Asn	TAT Tyr 10	AAT Asn	GAT Asp	CCT Pro	ATT Ile	GAT Asp 15	AAT Asn		48
 AAT Asn	AAT Asn	ATT Ile	ATT Ile 20	ATG Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	TTT Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr 30	GGG Gly	AGA Arg		96
TAT Tyr	TAT Tyr	AAA Lys 35	GCT Ala	TTT Phe	AAA Lys	ATC Ile	ACA Thr 40	GAT Asp	CGT Arg	ATT Ile	TGG Trp	ATA Ile 45	ATA Ile	CCG Pro	GAA Glu		144
AGA Arg	TAT Tyr 50	ACT Thr	TTT Phe	GGA Gly	TAT Tyr	AAA Lys 55	Pro	GAG Glu	Asp	Phe	AAT Asn -60	AAA Lys	AGT Ser	TCC Ser	GGT Gly		192
ATT Ile 65	Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT Tyr	TAT Tyr	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	•	240
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT Phe		288
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT Ile		336
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu		384
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn		432
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160		480
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly		528
ATA Ile	CAA Gln	Asn	His	Phe	GCA Ala	Ser	Arg	Glu	Gly	Phe	Gly	Gly	Ile	Met	CAA Gln		576
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu		624
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCÁ Pro		672
GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA Glu	CTT Leu	ATA Ile	CAT His	GTT Val 235	TTA Leu	CAT His	GGA Gly	Leu	TAT Tyr 240		720

GGC	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	GCA Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe	816
GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	ĠAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	.GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	Ser	AAG Lys	Glu	CAT His	Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	ĠAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536

GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys		1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	•	1632
ACA Thr 545	TTT Phe	CCT Pro	CTA Leu	GAT Asp	ATA Ile 550	AGA Arg	GAT Asp	ATA Ile	AGT Ser	TTA Leu 555	ACA Thr	TCT Ser	TCA Ser	TTT Phe	GAT Asp 560		1680
GAT Asp	GCA Ala	TTA Leu	TTA Leu	TTT Phe 565	TCT Ser	ÀAC Asn	AAA Lys	GTT Val	TAT Tyr 570	TCA Ser	TTT Phe	TTT	TCT Ser	ATG Met 575	GAT Asp		1728
TAT Tyr	ATT Ile	AAA Lys	ACT Thr 580	GCT Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	GAA Glu	GCA Ala	GGA Gly	TTA Leu	TTT Phe 590	GCA Ala	GGT Gly		1776
TGG Trp	GTG Val	AAA Lys 595	CAG Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	Phe	GTA Val	ATC Ile	GAA Glu	GCT Ala 605	AAT Asn	AAA Lys	ÄGC Ser		1824
AAT Asn	ACT Thr 610	ATG Met	GAT Asp	AAA Lys	ATT Ile	GCA Ala 615	GAT Asp	ÄTA Ile	TCT Ser	CTA Leu	ATT Ile 620	GTT Val	CCT Pro	TAT	ATA Ile		1872
GGA Gly 625	TTA Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	GGA Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	AAA Lys	GGA Gly	AAT Asn	TTT Phe	GAA Glu 640		1920
AAT Asn	GCT Ala	TTT	GAG Glu	ATT Ile 645	GCA Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	CTA Leu	CTA Leu	GAA Glu	TTT Phe	ATA Ile 655	CCA Pro		1968
GAA Glu	CTT Leu	TTA Leu	ATA Ile 660	CCT Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA Leu	GAA Glu	TCA Ser 670	TAT Tyr	ATT Ile		2016
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys		2064
AGA Arg	AAT Asn 690	GAA Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	ATG Met	TAC	GGA Gly	TTA Leu	ATA Ile 700	GTA Val	GCG Ala	CAA Gln	TGG Trp		2112
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720		2160
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr		2208
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp		2256
TTT Phe	AAT Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	AAA Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile		2304
GAT Asp	AAT Asn 770	Ile	AAT Asn	AAT Asn	TTT Phe	ATA Ile 775	AAT Asn	GGA Gly	TGT Cys	TCT Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met		2352

AAA Lys 785	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
TTG Leu	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	GAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
						AAA Lys 855										2574

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 858 amino acids
 - TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22: Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80 Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 185

155

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Glu Glu Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 250 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 455 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr . 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535

WO 98/07864 PCT/GB97/02273

- 100 -

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 555 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 695 Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 710 Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 760 Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 840 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser 850

- (2) INFORMATION FOR SEQ ID NO: 23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1644 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

- 101 -

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi)	SEQUENCE	DESCRIPTION:	SEO	ID	NO:	23:
------	----------	--------------	-----	----	-----	-----

ATG Met 1	CCA Pro	GTT Val	ACA Thr	ATA Ile 5	AAT Asn	AAT Asn	TTT Phe	AAT Asn	TAT Tyr 10	AAT Asn	GAT Asp	CCT Pro	ATT Ile	GAT Asp 15	AAT Asn	4	8
AAT Asn	AAT Asn	ATT Ile	ATT Ile 20	ATG Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	TTT Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr 30	GGG Gly	AGA Arg	9	6
TAT Tyr	TAT Tyr	AAA Lys 35	GCT Ala	TTT Phe	AAA Lys	ATC Ile	ACA Thr 40	GAT Asp	CGT	ATT Ile	TGG Trp	ATA Ile 45	ATA Ile	CCG Pro	GAA Glu	14	4
AGA Arg	TAT Tyr 50	ACT Thr	TTT Phe	GGA Gly	TAT Tyr	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	AAA Lys	AGT Ser	TCC Ser	GGT	. 19	2
ATT Ile 65	Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT Tyr	TAT	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	24	0
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT Phe	28	8
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT Ile	33	6
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	38	4
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn	43	2
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160	48	0
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	52	В
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	570	5
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624	4
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro	67:	2

GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAÁ Glu	CTT Leu	ATA Ile	CAT His	GTT Val 235	TTA Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240	7	20
GGC	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pró	ATT Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	Phe	7	68
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	GCA Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe	8	16
GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	8	64
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	9	12
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	9	60
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	10	80
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	10	56
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	11	.04
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	AȚA Ile	AAA Lys	11	.52
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	12	00
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	12	48
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	12	96
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	13	144
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	13	392
GAT Asp 465	Asp	TTA Leu	TCT	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	14	140
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	14	188

PCT/GB97/02273

- 103 -

		ATA Ile							1536
		GAT Asp							1584
		GAT Asp				_	 	 	1632
 	Pro		•						1644

- (2) INFORMATION FOR SEQ ID NO: 24:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 548 amino acids
 - (B) TYPE: amino acid (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 120 Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 185 Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 310 315 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 395 Ser Asp CLys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 445 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 455 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 535 540 Thr Phe Pro Leu 545

- 105 -

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:

 - (A) NAME/KEY: CDS
 (B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

	CAG Gln																48
	GAC Asp																.96
	AAG Lys																44
GAT Asp	ACA Thr 50	Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	1	.92
	AAG Lys															· 2	40
	AAC Asn																88
	ATT Ile															3	36
	GGA Gly															3	84
	ATT Ile 130															. 4	32
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	Asn	CTC Leu	GTA Val	ATC Ile	Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	4	80
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	. 5	28
CGT	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe		76

ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	. 768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	Asn	Ile	Val	CCT Pro 380	Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT	AAT Asn	GGT	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392

AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480		1440
ATT Ile	ACA Thr	TCT Ser	GAT Asp	ACT Thr 485	AAT Asn	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT	AGT Ser 495	TTA		1488
		Ile												Glu	CCT		1536
													Gly		TTA Leu		1584
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT	GAG Glu		1632
															GAA Glu 560	٠,	1680
						GCT Ala											1728
						TAT Tyr										•	1776
						GAG Glu											1824
						ACC Thr 615											1872
						ACT Thr									GCT Ala 640	. ·	1920
						TTA Leu											1968
						ATT Ile											2016
						TTT											2064
						ATA Ile 695											2112
						AAA Lys											2160
						CTA Leu											2208

GA G1	A AAT u Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CA G1	A TAT n Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
TI	A AGT u Ser 770	Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile	2352
AA As 78	T AAA n Lys 5	TTT	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
AT Il	C CCI e Pro	TAT	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GÄA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
GA As	T GCA p Ala	. TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CA G1	A GTA n Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
AT Il	A CCT e Pro 850	Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
	A TTT r Phe						TAA *									2616

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 (B) TYPE: amino acid

 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 70

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105

Arg	Gly	Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys
Val	Ile 130	Asp	Thr	Asn	Cys	Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Туг
Arg 145	Ser	Glu	Glu	Leu	Asn 150	Leu	Val	Ile	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160
Ile	Gln	Phe	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr
Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu
Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr.	Gly	Ile 235	Ala	Ile	Asn	bio	Asn 240
Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys
		275					280				Leu	285			
-	290					295					Ala 300				
305					310					315	Val				320
-				325					330		Ser			335	
	•		340					345			Ile		350		
		355					360				Lys	365			
	370					375					Pro 380				
385					390					395	Asn				400
				405					410		Asn			415	
•			420					425			Leu		430		
_		435					440				Lys	445			
Ala	Leu 450	Asn	Asp	Leu	Cys	Ile 455	Lys	Val	Asn	Asn	Trp 460	Asp	Leu	Phe	Phe

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 520 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 615 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 745 Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805

WO 98/07864 PCT/GB97/02273

- 111 -

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser

Thr Phe Thr Glu Tyr Ile Lys *

- (2) INFORMATION FOR SEQ ID NO: 27:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATGCCGGTTA	CCATCAACAA	CTTCAACTAC	AACGACCCGA	TCGACAACAA	CAACATCATC	60
ATGATGGAAC	CGCCGTTCGC	ACGTGGTACC	GGTCGTTACT	ACAAGGCTTI	CAAGATCACC	120
GACCGTATCT	GGATCATCCC	GGAACGTTAC	ACCTTCGGTT	ACAAACCTGA	GGACTTCAAC	180
AAGAGTAGCO	GGATTTTCAA	TCGTGACGTC	TGCGAGTACT	ATGATCCAGA	TTATCTGAAT	240
ACCAACGATA	AGAAGAACAT	ATTCCTTCAG	ACTATGATCA	AGTTATTTAA	TAGAATCAAA	300
TCAAAACCAT	TGGGTGAAAA	GTTATTAGAG	ATGATTATAA	ATGGTATACC	TTATCTTGGA	360
GATAGACGTG	TTCCACTCGA	AGAGTTTAAC	ACAAACATTG	CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA	ATCCAGGAGA	AGTGGAGCGA	AAAAAAGGTA	TTTTCGCAAA	TTTAATAATA	480
TTTGGACCTG	GGCCAGTTTT	AAATGAAAAT	GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA	GGGAAGGCTT	CGGGGGTATA	ATGCAAATGA	AGTTTTGCCC	AGAATATGTA	600
AGCGTATTTA	ATAATGTTCA	AGAAAACAAA	GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTTCAGATC	CAGCCTTGAT	ATTAATGCAT	GAACTTATAC	ATGTTTTACA	TGGATTATAT	720
GGCATTAAAG	TAGATGATTT	ACCAATTGTA	CCAAATGAAA	AAAAÄTTTTT	TATGCAATCT	780
ACAGATGCTA	TACAGGCAGA	AGAACTATAT	ACATTTGGAG	.GACAAGATCC	CAGCATCATA	840
ACTCCTTCTA	CGGATAAAAG	TATCTATGAT	AAAGTTTTGC	AAAATTTTAG	AGGGATAGTT	900
GATAGACTTA	ACAAGGTTTT	AGTTTGCATA	TCAGATCCTA	ACATTAATAT	TAATATATAT	960
AAAAATAAAT	TTAAAGATAA	ATATAAATTC	GTTGAAGATT	CTGAGGGAAA	ATATAGTATA	1020
GATGTAGAAA	GTTTTGATAA	ATTATATAAA	AGCTTAATGT	TTGGTTTTAC	AGAAACTAAT	1080
ATAGCAGAAA	ATTATAAAAT	AAAAACTAGA	GCTTCTTATT	TTAGTGATTC	CTTACCACCA	1140
GTAAAAATAA	AAAATTTATT	AGATAATGAA	ATCTATACTA	TAGAGGAAGG	GTTTAATATA	1200

WO 98/07864 PCT/GB97/02273

- 112 -

TCTGATAAAG	ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
ratgaagaaa	TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
AAAGCTCCAG	GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
AATAGTTTTT	CAGATGATTT	ATCŤAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
TATATAGAAA	ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ATACTGATTT	AATAAGTAAA	1500
ATAGAATTAC	CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
TATGAAAAAC	AACCCGCTAT	AAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
TTATACTCTC	AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	ATAAAAATAA	AAATTATTAA	2040
ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
AATGAGGGTA	TTAACCAÁGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATAA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
TCAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2574 base pairs
 (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: double (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ATGCCAGTTA	CAATAAATAA	TTTTAATTAT	AATGATCCTA	TTGATAATAA	TAATATTATT	60
ATGATGGAGC	CTCCATTTGC	GAGAGGTACG	GGGAGATATT	ATAAAGCTTT	TAAAATCACA	120
GATCGTATTT	GGATAATACC	GGAAAGATAT	ACTTTTGGAT	ATAAACCTGA	GGATTTTAAT	180
AAAAGTTCCG	GTATTTTTAA	TAGAGATGTT	TGTGAATATT	ATGATCCAGA	TTACTTAAAT	240

- 113 -

ACTAATGATA	AAAAGAATAT	ATTTTTACA	ACAATGATCA	AGTTATTTA	TAGAATCAAA	300
TCAAAACCAT	TGGGTGAAAA	GTTATTAGAG	ATGATTATAA	ATGGTATACO	TTATCTTGGA	360
GATAGACGTG	TTCCACTCGA	AGAGTTTAAC	ACAAACATTG	CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA	ATCCAGGAGA	AGTGGAGCGA	AAAAAAGGTA	TTTTCGCAAA	TTTAATAATA	480
TTTGGACCTG	GGCCAGTTTT	AAATGAAAAT	GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA	GGGAAGGCTT	CGGGGGTATA	ATGCAAATGA	AGTTTTGCCC	: AGAATATGTA	600
AGCGTATTTA	ATAATGTTCA	AGAAAACAAA	GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTTCAGATC	CAGCCTTGAT	ATTAATGCAT	GAACTCATCC	ACGTCCTCCA	CGGTCTCTAC	720
GGTATCAAAG	TAGACGACCT	CCCGATCGTC	CCGAACGAAA	AAAAATTCTT	CATGCAGAGC	780
ACCGACGCAA	TCCAGGCAGA	AGAACTCTAC	ACCTTCGGTG	GTCAGGACCC	GAGCATCATC	840
ACCCCGAGCA	CCGACAAAAG	CATCTACGAC	AAAGTCCTCC	AGAACTTCCG	TGGTATCGTC	900
GACCGTCTCA	ACAAAGTCCT	CGTCTGCATC	AGCGACCCGA	ACATCAACAT	CAACATCTAC	960
AAAAACAAAT	TCAAAGACAA	ATACAAATTC	GTCGAAGACA	GCGAAGGTAA	ATACAGCATC	1020
GACGTCGAGA	GCTTCGACAA	ACTCTACAAA	AGCCTCATGT	TCGGTTTCAC	CGAAACCAAC	1080
ATCGCAGAAA	ACTACAAAAT	CAAAACCCGT	GCAAGCTACT	TCAGCGACAG	CCTCCCGCCG	1140
GTCAAAATCA	AAAACCTCCT	CGACAACGAA	ATCTACACCA	TCGAAGAAGG	TTTCAACATC	1200
AGCGACAAAG	ACATGGAAAA	AGAATACCGT	GGTCAGAACA	AAGCAATCAA	CAAACAAGCT	1260
TACGAAGAAA	TCAGCAAAGA	ACACCTCGCA	GTCTACAAAA	TCCAGATGTG	CAAAAGCGTC	1320
AAAGCACCGG	GTATCTGCAT	CGACGTTGAC	AACGAAGACC	TCTTCTTCAT	CGCAGACAAA	1380
AACAGCTTCA	GCGACGACCT	CAGCAAAAAC	GAACGTATCG	AATACAACAC	CCAGAGCAAC	1440
TACATCGAAA	ACGACTTCCC	GATCAACGAA	CTCATCCTCG	ACACCGACCT	CATCAGCAAA	1500
ATCGAACTCC	CGAGCGAAAA	CACCGAAAGC	CTCACCGACT	TCAACGTTGA	CGTCCCGGTC	1560
TACGAAAAAC	AGCCGGCAAT	CAAAAAAATC	TTCACCGACG	AAAACACCAT	CTTCCAGTAC	1620
CTCTACAGCC	AGACCTTCCC	GCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	ATAAAAATAA	AAATTATTAAA	2040
ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280

WO 98/07864 PCT/GB97/02273

- 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
CATAT TTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
CAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

CLAIMS

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- 3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain H_N portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain H_N portion.
- 8. A polypeptide according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain β of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated $H_{\rm N}$ of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 17. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H_c of a botulinum toxin heavy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated H_c of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain H_N portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H_N portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
- 33. A fusion protein according to Claim 32 wherein said second polypeptide is

adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

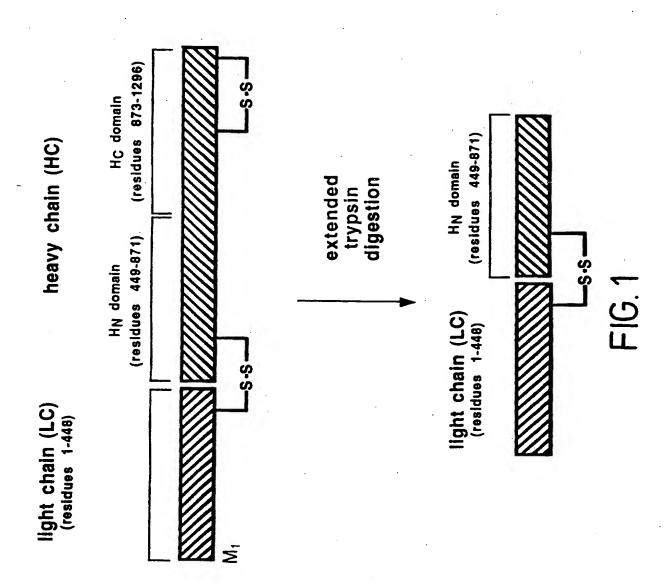
40 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.

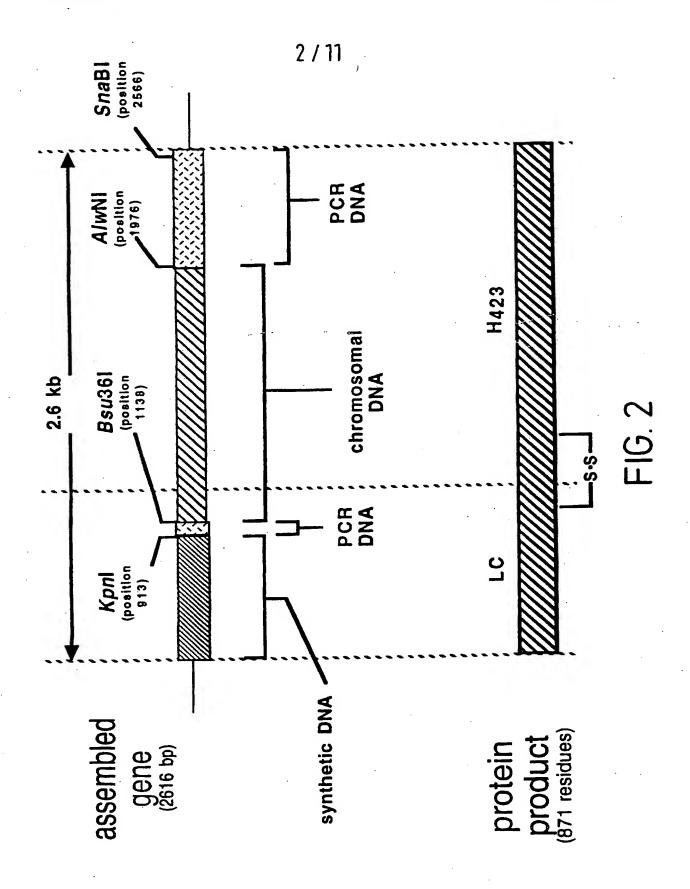
- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain H_N domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- 44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain H_N domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- 46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- 47. A DNA according to any of claims 40-46.
- 48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.
- 49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- 50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

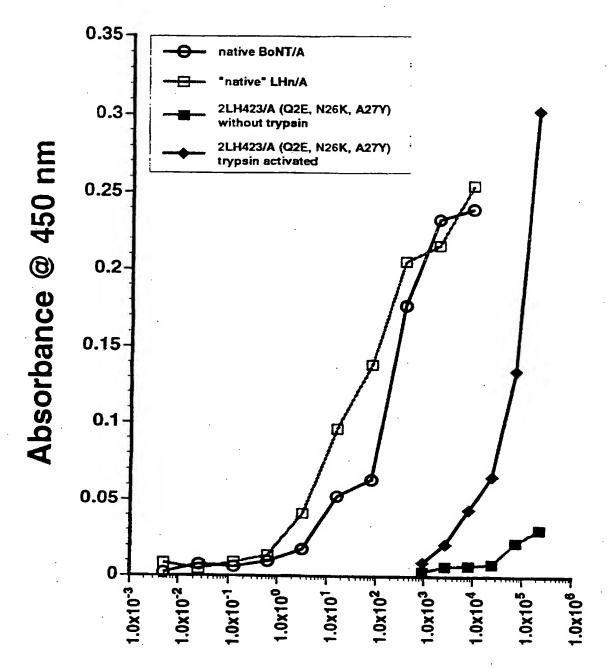
according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.

1/11







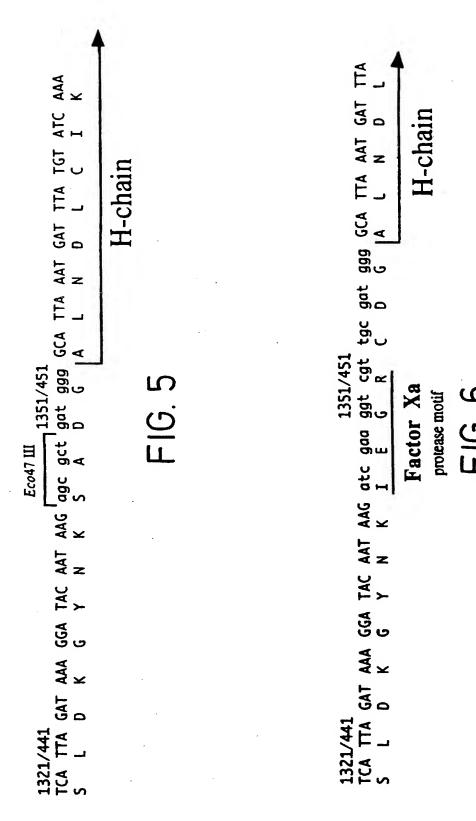
Protein concentration (ng/ml)

FIG. 3

		_	4/11			
NYKDPVNGVDIAYIKIPNAGQMQPV (Seq!D.2)		Q P V (Seq I.D. 4)	Q P V (Seq I.D. 6)	,		
<u> </u>		Ξ	브			
Sed		Sed	Sed			
<u></u>		<u>.</u>	. 9			
>		>	>	>	>	
<u>م</u>		<u>-</u>	~	. O .	0 P	
5		∑	∑	₹	∑	
2		0	0	∑	∑ ⊘	
ی		KIPKYGOM	∑ ⊘ ∪	ی	ی	
A		>-		<	V	
z		¥	<u>⊼</u>	z	IKIPNA	
	· ·	۵	۵	۵	۵.	
-	•	H	H	Д. У	Н	
¥		×	H X H	. x	× .	
" H		H	 	H		
>		>	>	. A	> '	
∢	<u> </u>	DIAY	DIAY	H	IA	
		-	6	<u>ہ</u>	_ _	
>	ш	>		>	>	
ف	<u></u>	ی	ی	6	5	
Z	ш	2	> 5 N	z	Z	
> .		>	д	>	>	
<u>α</u>	QKLL	۵.	<u>م</u>	٥	Δ.	
0	¥		×	0	0	
¥	0	∠ .	×	¥	×	
>	T R L	FNYKDPV	>	>	>	
Z.	~	Z	T.	Z	Z	
u.	—	ır.	ட	ī.	u.	
0	S	0	0	~	0	
~	⊢	×	~	~	~	
Z	X .	Z	Z	Z	2	
>	Ξ.	>	<i>></i>	<i>></i>	,	
MQFVNKQ	G S P G I H M T S	G T M E F V N K Q	G S M E F V N K	1 Q F V K	MPFVNK	
	<u>a</u>	₹ 54 × 1	Σ	∑ ਜ	<u>-</u>	
	S	7.	G S M 123			
	, 5 4	ر ی	5 4			
	L					
		٤	٤	Native BoNT/A, C. botulinum 2169 Thompson et al.1990	_	
		(Q ₂ E, N ₂₆ K, A ₂₇ Y)	₂ LH ₄₂₃ /A (Q ₂ E, N ₂₆ K, A ₂₇ Y)	Native BoNT/A, C. botulinum 2169 Thompson et al.19	Native BoNT/A, C. botulinum 62A Binz et al.1990	
ď	_	۶̈́	Ž,	E P	199 199	
153//	123/4	Ž	23/A	Bol Ilin Son	Bol Ilinu al.	
LH ₄₂₃ /A	23LH423/A	щ ²	2LH423/A (Q2E, N26	We We	ive ootu	
_	31	0	7, 0	Native BoNT/A, C. botulinum 21 Thompson et al	Native BoNT/A, C. botulinum 62 Binz el al.1990	
				2 0 1	- U	

ָר טַני

= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.



6/11

GAC AGG D R CAG ACA TAT Y 66A 6 AGG R ATG M CCT P GAG E TCT S 16T C 900 A 6TG V CT 6 AAG K AGG R ر 36 AGG R ATT I TTC F CAG AGT S CTA AGC S 2617/873
TTT ACT
F T
2677/893
GAT GCT
D A
2737/913
GGC TCC
G S
2797/933
TGT GAT
C D F) ACA T GAA E 6TG V TAT Y AGC S CTG L 6.66 R TCT S GCT A TTA L TCA S GAG E ACA T T F AAG K HA L GCT.)) | |)) (9 8 8 AGA R 999 AAG K))) CAA O 7 76C C AAC N GAG E AAG K CTC FF GAT D ٦ ٦ GAT D GTG V ACG T TAT Y 2587/863 TAC GTA Y V 2647/883 CCG GAG CCG GAG GGC TTT G F 2767/923 GGT ATC G I 2827/943 GCA CCC

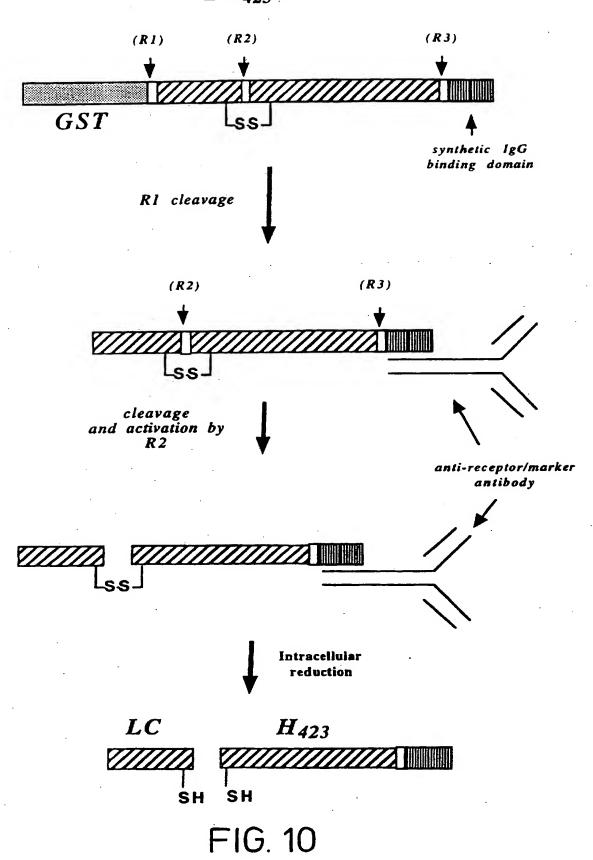
16.7

AAT N AGG R CAT H ACA T TCT S AAG K GAT D ATT I ATT I GAT D GAA E TCT S 2617/873 TTT ACT F T 2677/893 TAT CAA Y Q TCT ACA S T TCA GGC S G TTA L TA L ATA I TGA stop AGA R & 0 8 H L CAA. AGA R AAA K GAA GAT F > GAT D 2587/863 TAC GTA Y V 2647/883 TCT AAA S K 2707/903 ATT AAG

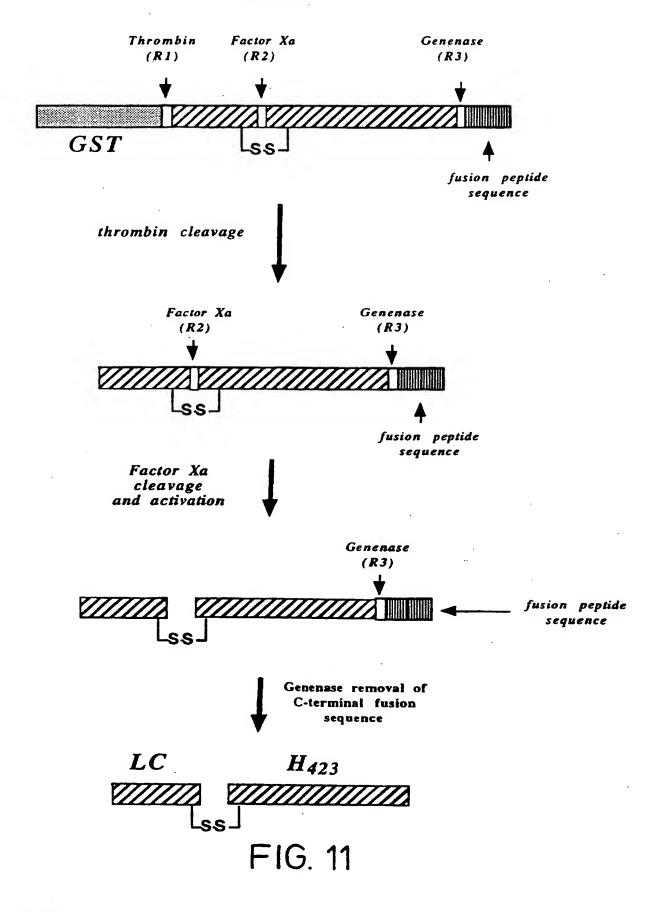
=1G.8

2587/863	3	1	3	Ş	Ē	Ě	171	V	2617/873	/873 ACT	*	, ⊢ +	Į	7	T .			AAT
<u>5</u> >	- CA -	A A Z	\$ 5 0	A	ر د ج	ا - ا	ر د د	T & T	- 1	- - - -	¥ Эш	- ₹ - >-		2 X X	S	ی و	ב ב	(z
/88	ന		,						2677,	/893								
5	GGT	S	ST	RAT	TAT	939	CAA	CAC	GAT	GAA	J J	GTA	SAC	AAC	AAA			AAA
م	G	4	4	I	_	4	0	I	۵	ш	⋖	>	۵	z	¥			¥
7/90	3				1		,		2737,	/913								
S	CAA	AAC	93	ĭ	TAT	GAG	ATC	TA	CAT	TTA	5	AAC	TTA	AAC	GAA			S S
0	0	z	«	ı.	>	ш	ı	_	x	_	۵	z	_	z	w			œ
7/92	, 								2797/	/933								,
9	<u> </u>	ATC	SA	AGT	TA T	AAA	GAT	GAC	5	AGC	CAA	AGC	SCT CT	AAC	E			GAA
4	u.	H	0	S	_	¥	۵	۵	۵	S	0	S	4	z				ш
7/94	~		•						2857/	/953								
AAA	AAG	CTA	AAT	GAT	L ₃	CAG	929	900	AAA	GTA	GAC	AAC	AAA	110	AAC			CAA
×	¥	_	z	۵	4	0	A	۵	×	>	_	z	×	u.	z			0
36/2	m					•			2917/	/973								
AAC	939	ĭ	TAT	GAG	ATC	T T	S	TA	5	AAC	TTA	AAC	GAA	GAA	CAA			ပ္ပ ပြွ
z	V	u.	>	ш	_		ェ		۵.	z	_	z	u	ய	0			4
7/98	~								2977/	/993								
ATC	CAA	AGT	¥	AAA	GAT	SAC	SC	AGC	CAA	AGC	SCT CT	AAC	E	TTA	SCA SCA	GAA		AAA
Н	0	S	_	¥	۵	۵	۵	S	0	S	⋖	z	_	_	Ø			¥
7/100	3 3`								3037/	/1013								
CTA	AAT	GAT	ST	S	ဗ္ဗ	ဗ္ဗ	AAA	GTA	SAC	TAG								
_	z	۵	4	0	4	۵	¥	>	D	*								
							П	<u></u>	σ									

8/11 LH₄₂₃/A



$LH_{423}/A^{9/11}$



10/11

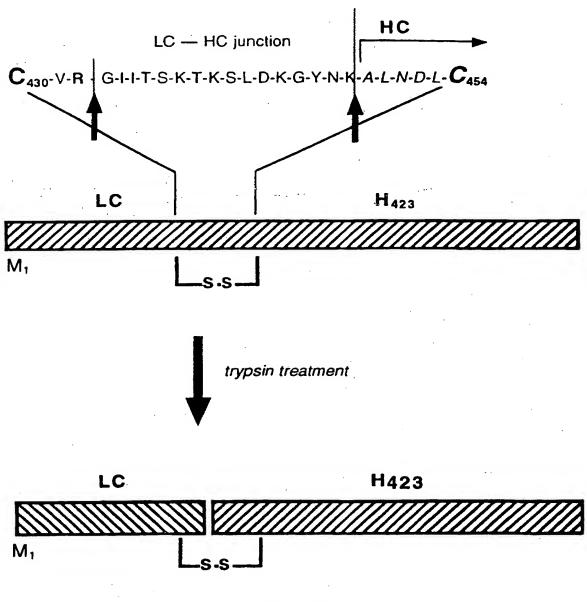
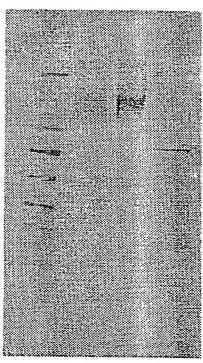


FIG. 12

11 / 11

Panel A. 1 2 3 4



Panel B. 1 2 3 4

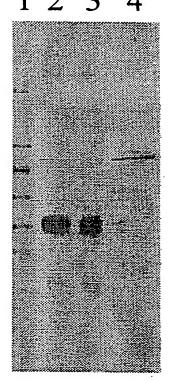


FIG. 13

INTERNATIONAL SEARCH REPORT

Inten onal Application No PCT/GB 97/02273

A. CLASSIFICATION OF SUBJECT MATTER
1PC 6 C12N15/31 C12N1/21 C07K14/33 C12P21/02 A61K38/16 A61K39/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12P A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 12802 A (OPHIDIAN PHARM INC; WILLIAMS JAMES A (US); PADHYE NISHA V (US); KI) 2 May 1996 see the whole document	1-52
X	KURAZONO H ET AL: "Minimal essential *domains* specifying toxicity of the *light* *chains* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II	1-52

X Further documents are fisted in the continuation of box C.	X Patent family members are listed in annex.
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but tater than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family
Date of the actual completion of the international search	Date of mailing of the international search report
9 December 1997	3 0. ot 98
Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentiaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Fax: (+31-70) 340-3016	- Hillenbrand, G

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter. .onal Application No
PCT/GB 97/02273

		PCT/GB 97/0	22/3
	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	l Pa	evant to claim No.
ategory *	Citation of document, with indication, where appropriate, of the relevant passages		evant to claim No.
(LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document		1
A	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document		1,26,35
		(1)	
		!	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Intel. .onal Application No
PCT/GB 97/02273

Patent document cited in search report WO 9612802 A 02-05-96 US 5601823 A 11-02-97 US 5196193 A 23-03-93 AU 3968395 A EP 0796326 A 24-09-97 FI 971732 A 23-06-97 NO 971868 A 24-06-97 PL 320214 A 15-09-97 ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 23-06-94 US 5466672 A 14-11-95 US 5599539 A 04-02-97	
US 5196193 A 23-03-93 AU 3968395 A 15-05-96 EP 0796326 A 24-09-97 FI 971732 A 23-06-97 NO 971868 A 24-06-97 PL 320214 A 15-09-97 ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 WO 9413264 A 23-06-94 US 5466672 A 14-11-95	
AU 3968395 A 15-05-96 EP 0796326 A 24-09-97 FI 971732 A 23-06-97 NO 971868 A 24-06-97 PL 320214 A 15-09-97 ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 WO 9413264 A 23-06-94 US 5466672 A 14-11-95	
EP 0796326 A 24-09-97 FI 971732 A 23-06-97 NO 971868 A 24-06-97 PL 320214 A 15-09-97 ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 WO 9413264 A 23-06-94 US 5466672 A 14-11-95	
FI 971732 A 23-06-97 NO 971868 A 24-06-97 PL 320214 A 15-09-97 ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 WO 9413264 A 23-06-94 US 5466672 A 14-11-95	
NO 971868 A 24-06-97 PL 320214 A 15-09-97 ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 WO 9413264 A 23-06-94 US 5466672 A 14-11-95	
PL 320214 A 15-09-97 ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 W0 9413264 A 23-06-94 US 5466672 A 14-11-95	
ZA 9508990 A 15-05-96 AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 W0 9413264 A 23-06-94 US 5466672 A 14-11-95	
AU 6653894 A 04-07-94 CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 W0 9413264 A 23-06-94 US 5466672 A 14-11-95	
CA 2150935 A 23-06-94 EP 0671902 A 20-09-95 W0 9413264 A 23-06-94 US 5466672 A 14-11-95	
EP 0671902 A 20-09-95 W0 9413264 A 23-06-94 US 5466672 A 14-11-95	
WO 9413264 A 23-06-94 US 5466672 A 14-11-95	•
US 5466672 A 14-11-95	
US 5599539 A 04-02-97	
AU 638786 B 08-07-93	
AU 6895191 A 31-05-91	
EP 0498854 A 19-08-92	
WO 9106306 A 16-05-91	
US 5443976 A 22-08-95	
US 5340923 A 23-08-94	